An Introduction to Higher Mathematics

Patrick Keef
David Guichard

This text was initially written by Patrick Keef and modified by David Guichard.

This copy of the text was compiled from source at 17:50 on 10/14/2021.

We will be glad to receive corrections and suggestions for improvement at guichard@whitman.edu.

Brief Contents

Contents

2

Proofs 35

3

Number Theory 53

4

Functions 89

5

Relations 117

Bibliography 131

Index 133

1

Logic

Although mathematical ability and opinions about mathematics vary widely, even among educated people, there is certainly widespread agreement that mathematics is *logical*. Indeed, properly conceived, this may be one of the most important defining properties of mathematics.

Logical thought and logical arguments are not easy to come by, nor is it always clear whether a given argument is logical (that is, logically correct). Logic itself deserves study; the right tools and concepts can make logical argument easier to discover and to discern. In fact, logic is a major and active area of mathematics; for our purposes, a brief introduction will give us the means to investigate more traditional mathematics with confidence.

1.1 Logical Operations

Mathematics typically involves combining true (or hypothetically true) statements in various ways to produce (or prove) new true statements. We begin by clarifying some of these fundamental ideas.

By a **sentence** we mean a statement that has a definite **truth value**, true (T) or false (F)—for example,

"In 1492 Columbus sailed the ocean blue." (T)

"Napoleon won the battle of Waterloo." (F)

More generally, by a **formula** we mean a statement, possibly involving some variables, which is either true or false whenever we assign particular values to each of the variables.

We will use capital letters to designate formulas. If the truth of a formula depends on the values of, say, x, y and z, we will use notation like $P(x, y, z)$ to denote the formula.

EXAMPLE 1.1.1 If $P(x, y)$ is "$x^2 + y = 12$", then $P(2, 8)$ and $P(3, 3)$ are true, while $P(1, 4)$ and $P(0, 6)$ are false. If $Q(x, y, z)$ is "$x+y < z$", then $Q(1, 2, 4)$ is true and $Q(2, 3, 4)$ is false. □

Whether a sentence is true or false usually depends on what we are talking about—exactly the same sentence may be true or false depending on the context; for example, the formula $x|y$ means 'x divides y'. That is, $x|y$ if there is some z so that $y = x \cdot z$. Now, is it true that $3|2$? It depends: if we are talking about integers, the answer is no; if we are talking about rational numbers, the answer is yes, because $2 = 3 \cdot (2/3)$. (Of course, if $x \neq 0$ and y are *any* rational numbers then $x|y$, so that it is not a very useful notion. In normal usage, the appearance of the formula "$x|y$" *implies* that x and y are integers.)

The **universe of discourse** for a particular branch of mathematics is a set that contains everything of interest for that subject. When we are studying mathematical formulas like 'x divides y' the variables are assumed to take values in whatever universe of discourse is appropriate for the particular subject. The universe of discourse usually is clear from the discussion, but occasionally we will need to identify it explicitly for clarity. The universe of discourse is usually denoted by U.

Complicated sentences and formulas are put together from simpler ones, using a small number of **logical operations**. Just a handful of these operations will let us say everything we need to say in mathematics.

If P is a formula, then "not P" is another formula, which we write symbolically as $\neg P$. Of course, $\neg P$ is false if P is true and vice versa—for example,

"6 is not a prime number" or "It is not true that 6 is prime" or "¬(6 is prime)" (T)

"Ronald Reagan was not a president." (F)

Suppose that P and Q are formulas. Then "P and Q" is a formula written symbolically as $P \wedge Q$, called the **conjunction** of P and Q. For $P \wedge Q$ to be true both P and Q must be true, otherwise it is false—for example,

"$5 = 6$ and $7 = 8$." (F)

"Seattle is in Washington and Boise is in Idaho." (T)

"Tolstoy was Russian and Dickens was French." (F)

If P and Q are formulas, then the formula "P or Q" is written symbolically as $P \vee Q$, called the **disjunction** of P and Q. It is important to note that this is an *inclusive* or, that is, "either or both". So if P, Q or *both* P and Q are true, so is $P \vee Q$. The only way $P \vee Q$ can be false is if both P and Q are false—for example,

"Washington is in Canada or London is in England." (T)

"$5 < 7$ or $8 < 10$." (T)

"Lenin was Spanish or Ghandi was Italian." (F)

If P and Q are formulas, then "if P then Q" or "P implies Q" is written $P \Rightarrow Q$, using the **conditional** symbol, $\Rightarrow$. It is not obvious (at least, not to most people) under what circumstances $P \Rightarrow Q$ should be true. In part this is because "if...then" is used in more than one way in ordinary English, yet we need to fix a rule that will let us know precisely when $P \Rightarrow Q$ is true. Certainly, if P is true and Q is false, P cannot imply Q, so $P \Rightarrow Q$ is false in this case. To help us with the other cases, consider the following statement:

"If x is less than 2 then x is less than 4."

This statement should be true regardless of the value of x (assuming that the universe of discourse is something familiar, like the integers). If x is 1, it evaluates to T $\Rightarrow$ T, if x is 3, it becomes F $\Rightarrow$ T, and if x is 5, it becomes F $\Rightarrow$ F. So it appears that $P \Rightarrow Q$ is true unless P is true and Q is false. This is the rule that we adopt.

Finally, the **biconditional**, written $\Leftrightarrow$, corresponds to the phrase "if and only if" or "iff" for short. So $P \Leftrightarrow Q$ is true when both P and Q have the same truth value, otherwise it is false.

EXAMPLE 1.1.2 Suppose $P(x, y)$ is "$x + y = 2$" and $Q(x, y)$ is "$xy > 1$." Then when $x = 1$ and $y = 1$, $\neg P(x, y)$, $P(x, y) \wedge Q(x, y)$, $P(x, y) \vee Q(x, y)$, $P(x, y) \Rightarrow Q(x, y)$ and $P(x, y) \Leftrightarrow Q(x, y)$ have truth values F, F, T, F, F, respectively, and when $x = 2$ and $y = 3$ they have truth values T, F, T, T, F, respectively. □

Using the operations $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$, we can construct **compound** expressions such as

$$(P \wedge (\neg Q)) \Rightarrow ((\neg R) \vee ((\neg P) \wedge Q)).$$

As this example illustrates, it is sometimes necessary to include many parentheses to make the grouping of terms in a formula clear. Just as in algebra, where multiplication takes precedence over addition, we can eliminate some parentheses by agreeing on a particular order in which logical operations are performed. We will apply the operations in this order, from first to last: $\neg$, $\wedge$, $\vee$, $\Rightarrow$ and $\Leftrightarrow$. So

$$A \Rightarrow B \vee C \wedge \neg D$$

is short for

$$A \Rightarrow (B \vee (C \wedge (\neg D))).$$

Just as in algebra, it often is wise to include some extra parentheses to make certain the intended meaning is clear.

Much of the information we have discussed can be summarized in **truth tables**. For example, the truth table for $\neg P$ is:

P	$\neg P$
T	F
F	T

This table has two rows because there are only two possibilities for the truth value of P. The other logical operations use two variables, so they require 4 rows in their truth tables.

P	Q	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$	$P \Leftrightarrow Q$
T	T	T	T	T	T
F	T	F	T	T	F
T	F	F	T	F	F
F	F	F	F	T	T

Any compound expression has a truth table. If there are n simple (that is, not compound) formulas in the expression there will be 2^n rows in the table, because there are this many different ways to assign T's and F's to the n simple formulas in the compound expression. The truth table for $(P \wedge Q) \vee \neg R$ is

P	Q	R	$P \wedge Q$	$\neg R$	$(P \wedge Q) \vee \neg R$
T	T	T	T	F	T
F	T	T	F	F	F
T	F	T	F	F	F
F	F	T	F	F	F
T	T	F	T	T	T
F	T	F	F	T	T
T	F	F	F	T	T
F	F	F	F	T	T

Observe how the inclusion of intermediate steps makes the table easier to calculate and read.

A **tautology** is a logical expression that always evaluates to T, that is, the last column of its truth table consists of nothing but T's. A tautology is sometimes said to be **valid**; although "valid" is used in other contexts as well, this should cause no confusion. For example, $(P \wedge Q) \vee P \Leftrightarrow P$ is a tautology, since its truth table is:

P	Q	$P \wedge Q$	$(P \wedge Q) \vee P$	$(P \wedge Q) \vee P \Leftrightarrow P$
T	T	T	T	T
F	T	F	F	T
T	F	F	T	T
F	F	F	F	T

We list a few important tautologies in the following theorem.

THEOREM 1.1.3 The following are valid:

a) $P \Leftrightarrow \neg\neg P$

b) $P \vee Q \Leftrightarrow Q \vee P$

c) $P \wedge Q \Leftrightarrow Q \wedge P$

d) $(P \wedge Q) \wedge R \Leftrightarrow P \wedge (Q \wedge R)$

e) $(P \vee Q) \vee R \Leftrightarrow P \vee (Q \vee R)$

f) $P \wedge (Q \vee R) \Leftrightarrow (P \wedge Q) \vee (P \wedge R)$

g) $P \vee (Q \wedge R) \Leftrightarrow (P \vee Q) \wedge (P \vee R)$

h) $(P \Rightarrow Q) \Leftrightarrow (\neg P \vee Q)$

i) $P \Rightarrow (P \vee Q)$

j) $P \wedge Q \Rightarrow Q$

k) $(P \Leftrightarrow Q) \Leftrightarrow ((P \Rightarrow Q) \wedge (Q \Rightarrow P))$

l) $(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P)$

Proof. The proofs are left as exercises. ■

Observe that (b) and (c) are commutative laws, (d) and (e) are associative laws and (f) and (g) say that $\wedge$ and $\vee$ distribute over each other. This suggests that there is a form of algebra for logical expressions similar to the algebra for numerical expressions. This subject is called **Boolean Algebra**, and has many uses, particularly in computer science.

If two formulas always take on the same truth value no matter what elements from the universe of discourse we substitute for the various variables, then we say they are **equivalent**. The value of equivalent formulas is that they say the same thing. It is always a valid step in a proof to replace some formula by an equivalent one. In addition, many tautologies contain important ideas for constructing proofs. For example, (k) says that if you wish to show that $P \Leftrightarrow Q$, it is possible (and often advisable) to break the proof into two parts, one proving the implication $P \Rightarrow Q$ and the second proving the **converse**, $Q \Rightarrow P$.

In reading through theorem 1.1.3 you may have noticed that $\wedge$ and $\vee$ satisfy many similar properties. These are called "dual" notions—for any property of one, there is a nearly identical property that the other satisfies, with the instances of the two operations interchanged. This often means that when we prove a result involving one notion, we get the corresponding result for its dual with no additional work.

George Boole. Boole (1815–1864) had only a common school education, though he learned Greek and Latin on his own. He began his career as an elementary school teacher, but decided that he needed to know more about mathematics, so he began studying mathematics, as well as the languages he needed to read contemporary literature in mathematics. In 1847, he published a short book, *The Mathematical Analysis of Logic*, which may fairly be said to have founded the study of mathematical logic. The key contribution of the work was in redefining 'mathematics' to mean not simply the 'study of number and magnitude,' but the study of symbols and their manipulation according to certain rules. The importance of this level of abstraction for the future of mathematics would be difficult to overstate. Probably on the strength of this work, he moved into a position at Queens College in Cork.

In *Investigation of the Laws of Thought*, published in 1854, Boole established a real formal logic, developing what today is called Boolean Algebra, or sometimes the **algebra of sets**. He used the symbols for addition and multiplication as operators, but in a wholly abstract sense. Today these symbols are still sometimes used in Boolean algebra, though the symbols '$\wedge$' and '$\vee$', and '$\cap$' and '$\cup$', are also used. Boole applied algebraic manipulation to the process of reasoning. Here's a simple example of the sort of manipulation he did: The equation $xy = x$ (which today might be written $x \wedge y = x$ or $x \cap y = x$) means that 'all things that satisfy x satisfy y,' or in our terms, $x \Rightarrow y$. If also $yz = y$ (that is, $y \Rightarrow z$) then substituting $y = yz$ into $xy = x$ gives $x(yz) = x$ or $(xy)z = x$. Replacing xy by x, we get $xz = x$, or $x \Rightarrow z$. This simple example of logical reasoning is used over and over in mathematics.

In 1859, Boole wrote *Treatise on Differential Equations*, in which he introduced the algebra of differential operators. Using D to stand for 'the derivative of,' the differential equation $ay'' + by' + cy = 0$ may be written as $aD^2(y) + bD(y) + cy = 0$, or as $(aD^2 + bD + c)y = 0$. Remarkably, the solution to $aD^2 + bD + c = 0$, treating D as a *number*, provides information about the solutions to the differential equation.

The information here is taken from *A History of Mathematics,* by Carl B. Boyer, New York: John Wiley and Sons, 1968. For more information, see *Lectures on Ten British Mathematicians*, by Alexander Macfarlane, New York: John Wiley & Sons, 1916.

Exercises 1.1.

1. Construct truth tables for the following logical expressions:

a) $(P \wedge Q) \vee \neg P$

b) $P \Rightarrow (Q \wedge P)$

c) $(P \wedge Q) \Leftrightarrow (P \vee \neg R)$

d) $\neg P \Rightarrow \neg(Q \vee R)$

2. Verify the tautologies in Theorem 1.1.3.

3. Suppose $P(x, y)$ is the formula "$x + y = 4$" and $Q(x, y)$ is the formula "$x < y$". Find the truth values for the formulas

$$P(x,y) \wedge Q(x,y), \quad \neg P(x,y) \vee Q(x,y),$$
$$P(x,y) \Rightarrow \neg Q(x,y), \quad \neg(P(x,y) \Leftrightarrow Q(x,y)),$$

using the values:

a) $x = 1, y = 3$ **c)** $x = 1, y = 2$

b) $x = 3, y = 1$ **d)** $x = 2, y = 1$

4. **a)** Find truth tables for

$$P \wedge (\neg Q) \wedge R, \qquad (\neg P) \wedge Q \wedge (\neg R)$$

b) Use these to find a truth table for

$$(P \wedge (\neg Q) \wedge R) \vee ((\neg P) \wedge Q \wedge (\neg R))$$

c) Use the method suggested by parts (a) and (b) to find a formula with the following truth table.

P	Q	R	???
T	T	T	T
F	T	T	F
T	F	T	F
F	F	T	F
T	T	F	T
F	T	F	T
T	F	F	F
F	F	F	F

d) Use the method suggested by parts (a)-(c) to explain why any list of 2^n T's and F's is the final column of the truth table for some formula with n letters.

5. If $P_1, P_2, \ldots, P_n$ is a list of n formulas, at most how many compound formulas using this list can be constructed, no two of which are equivalent?

1.2 Quantifiers

Recall that a formula is a statement whose truth value may depend on the values of some variables. For example,

"$x \leq 5 \wedge x > 3$"

is true for $x = 4$ and false for $x = 6$. Compare this with the statement

"For every x, $x \leq 5 \wedge x > 3$,"

which is definitely false and the statement

"There exists an x such that $x \leq 5 \wedge x > 3$,"

which is definitely true. The phrase "for every x" (sometimes "for all x") is called a **universal quantifier** and is denoted by $\forall x$. The phrase "there exists an x such that" is called an **existential quantifier** and is denoted by $\exists x$. A formula that contains variables is not simply true or false unless each of these variables is **bound** by a quantifier. If a variable is not bound the truth of the formula is contingent on the value assigned to the variable from the universe of discourse.

We were careful in section 1.1 to define the truth values of compound statements precisely. We do the same for $\forall x\, P(x)$ and $\exists x\, P(x)$, though the intended meanings of these are clear.

The Universal Quantifier

A sentence $\forall x\, P(x)$ is true if and only if $P(x)$ is true no matter what value (from the universe of discourse) is substituted for x.

EXAMPLE 1.2.1

- $\forall x\ (x^2 \geq 0)$, i.e., "the square of any number is not negative."
- $\forall x\, \forall y\ (x + y = y + x)$, i.e., the commutative law of addition.
- $\forall x\, \forall y\, \forall z\ ((x + y) + z = x + (y + z))$, i.e., the associative law of addition.

□

The "all" form. The universal quantifier is frequently encountered in the following context:

$$\forall x\ (P(x) \Rightarrow Q(x)),$$

which may be read, "All x satisfying $P(x)$ also satisfy $Q(x)$." Parentheses are crucial here; be sure you understand the difference between the "all" form and $\forall x\, P(x) \Rightarrow \forall x\, Q(x)$ and $(\forall x\, P(x)) \Rightarrow Q(x)$.

The latter formula might also be written as $\forall x\, P(x) \Rightarrow Q(x)$, which is to say that the universal quantifier has higher precedence than the conditional; to avoid misunderstanding, it is best to include the parentheses. The meaning of this formula might not be clear at first. The x in $P(x)$ is bound by the universal quantifier, but the x in $Q(x)$ *is not*. The formula $(\forall x\, P(x)) \Rightarrow Q(x)$ has the same meaning as $(\forall x\, P(x)) \Rightarrow Q(y)$, and its truth depends on the value assigned to the variable in $Q(\cdot)$.

EXAMPLE 1.2.2

- $\forall x$ (x is a square $\Rightarrow$ x is a rectangle), i.e., "all squares are rectangles."
- $\forall x$ (x lives in Walla Walla $\Rightarrow$ x lives in Washington), i.e., "every person who lives in Walla Walla lives in Washington."

□

This construction sometimes is used to express a mathematical sentence of the form "if this, then that," with an "understood" quantifier.

EXAMPLE 1.2.3

- If we say, "if x is negative, so is its cube," we usually mean "every negative x has a negative cube." This should be written symbolically as $\forall x\ ((x < 0) \Rightarrow (x^3 < 0))$.
- "If two numbers have the same square, then they have the same absolute value" should be written as $\forall x\, \forall y\ ((x^2 = y^2) \Rightarrow (|x| = |y|))$.
- "If $x = y$, then $x+z = y+z$" should be written as $\forall x\, \forall y\, \forall z\ ((x = y) \Rightarrow (x+z = y+z))$.

□

If S is a set, the sentence "every x in S satisfies $P(x)$" is written formally as

$$\forall x\ ((x \in S) \Rightarrow P(x))$$

For clarity and brevity, this is usually written $\forall x \in S\,(P(x))$. To understand and manipulate the formula $\forall x \in S\,(P(x))$ properly, you will sometimes need to "unabbreviate" it, rewriting it as $\forall x\ ((x \in S) \Rightarrow P(x))$.

EXAMPLE 1.2.4

- $\forall x \in [0,1]\ (\sqrt{x} \geq x)$ stands for $\forall x\ (x \in [0,1] \Rightarrow \sqrt{x} \geq x)$.
- $\forall x < 0\ (|x| = -x)$ stands for $\forall x\ (x < 0 \Rightarrow |x| = -x)$.

□

The Existential Quantifier

A sentence $\exists x\, P(x)$ is true if and only if there is at least one value of x (from the universe of discourse) that makes $P(x)$ true.

EXAMPLE 1.2.5

- $\exists x\ (x \geq x^2)$ is true since $x = 0$ is a solution. There are many others.
- $\exists x\, \exists y\ (x^2 + y^2 = 2xy)$ is true since $x = y = 1$ is one of many solutions.

□

The "some" form. The existential quantifier is frequently encountered in the following context:

$$\exists x\ (P(x) \wedge Q(x)),$$

which may be read, "Some x satisfying $P(x)$ also satisfies $Q(x)$."

EXAMPLE 1.2.6

- $\exists x\,(x$ is a professor $\wedge$ x is a republican), i.e., "some professor is a republican."

- $\exists x\,(x$ is a prime number $\wedge\ x$ is even), i.e., "some prime number is even."

□

It may at first seem that "Some x satisfying $P(x)$ satisfies $Q(x)$" should be translated as

$$\exists x\ (P(x) \Rightarrow Q(x)),$$

like the universal quantifier. To see why this does not work, suppose $P(x) =$ "x is an apple" and $Q(x) =$ "x is an orange." The sentence "some apples are oranges" is certainly false, but

$$\exists x\ (P(x) \Rightarrow Q(x))$$

is true. To see this suppose x_0 is some particular orange. Then $P(x_0) \Rightarrow Q(x_0)$ evaluates to F $\Rightarrow$ T, which is T, and the existential quantifier is satisfied.

We use abbreviations of the "some" form much like those for the "all" form.

EXAMPLE 1.2.7

- $\exists x < 0\ (x^2 = 1)$ stands for $\exists x\ ((x < 0) \wedge (x^2 = 1))$
- $\exists x \in [0, 1]\ (2x^2 + x = 1)$ stands for $\exists x\ ((x \in [0, 1]) \wedge (2x^2 + x = 1))$ □

If $\forall$ corresponds to "all" and $\exists$ corresponds to "some" do we need a third quantifier to correspond to "none"? As the following shows, this is not necessary:

EXAMPLE 1.2.8

- "No democrats are republicans," can be written $\forall x$ (x is a democrat $\Rightarrow$ x is not a republican).
- "No triangles are rectangles," can be written $\forall x$ (x is a triangle $\Rightarrow$ x is not a rectangle).

□

In general, the statement "no x satisfying $P(x)$ satisfies $Q(x)$" can be written

$$\forall x\ (P(x) \Rightarrow \neg Q(x)).$$

(You may wonder why we do not use $\neg\exists x\,(P(x) \wedge Q(x))$. In fact, we could—it is equivalent to $\forall x\ (P(x) \Rightarrow \neg Q(x))$.)

Exercises 1.2.

In these problems, assume the universe of discourse is the real numbers.

1. Express the following as formulas involving quantifiers:

a) Any number raised to the fourth power is non-negative.

b) Some number raised to the third power is negative.

c) The sine of an angle is always between $+1$ and -1.

d) The secant of an angle is never strictly between $+1$ and -1.

2. Suppose X and Y are sets. Express the following as formulas involving quantifiers.

a) Every element of X is an element of Y.

b) Some element of X is an element of Y.

c) Some element of X is not an element of Y.

d) No element of X is an element of Y.

3. Recall (from calculus) that a function f is **increasing** if

$$\forall a\ \forall b\ (a < b \Rightarrow f(a) < f(b))$$

Express the following definitions as formulas involving quantifiers:

a) f is **decreasing**.

b) f is **constant**.

c) f has a **zero**.

4. Express the following laws symbolically:

a) the commutative law of multiplication

b) the associative law of multiplication

c) the distributive law

5. Are the following sentences true or false?

a) $\forall x\ \forall y\ (x < y \Rightarrow x^2 < y^2)$

b) $\forall x\ \forall y\ \forall z \neq 0\ (xz = yz \Rightarrow x = y)$

c) $\exists x < 0\ \exists y < 0\ (x^2 + xy + y^2 = 3)$

d) $\exists x\ \exists y\ \exists z\ (x^2 + y^2 + z^2 = 2xy - 2 + 2z)$

6. Suppose $P(x)$ and $Q(y)$ are formulas.

a) Is $\forall x\ \forall y\ (P(x) \Rightarrow Q(y))$ equivalent to $\forall x(P(x)) \Rightarrow \forall y(Q(y))$?

b) Is $\exists x\ \exists y\ (P(x) \wedge Q(y))$ equivalent to $\exists x(P(x)) \wedge \exists y(Q(y))$?

1.3 De Morgan's Laws

If P is some sentence or formula, then $\neg P$ is called the **denial** of P. The ability to manipulate the denial of a formula accurately is critical to understanding mathematical arguments. The following tautologies are referred to as De Morgan's laws:

$$\neg(P \vee Q) \Leftrightarrow (\neg P \wedge \neg Q)$$
$$\neg(P \wedge Q) \Leftrightarrow (\neg P \vee \neg Q)$$

These are easy to verify using truth tables, but with a little thought, they are not hard to understand directly. The first says that the only way that $P \vee Q$ can fail to be true is if both P and Q fail to be true. For example, the statements "I don't like chocolate or vanilla" and "I do not like chocolate and I do not like vanilla" clearly express the same thought. For a more mathematical example of the second tautology, consider "x is not

between 2 and 3." This can be written symbolically as $\neg((2 < x) \wedge (x < 3))$, and clearly is equivalent to $\neg(2 < x) \vee \neg(x < 3)$, that is, $(x \leq 2) \vee (3 \leq x)$.

We can also use De Morgan's laws to simplify the denial of $P \Rightarrow Q$:

$$\begin{aligned} \neg(P \Rightarrow Q) &\Leftrightarrow \neg(\neg P \vee Q) \\ &\Leftrightarrow (\neg\neg P) \wedge (\neg Q) \\ &\Leftrightarrow P \wedge \neg Q \end{aligned}$$

so the denial of $P \Rightarrow Q$ is $P \wedge \neg Q$. In other words, it is not the case that P implies Q if and only if P is true and Q is false. Of course, this agrees with the truth table for $P \Rightarrow Q$ that we have already seen.

There are versions of De Morgan's laws for quantifiers:

$$\begin{aligned} \neg\forall x\, P(x) &\Leftrightarrow \exists x\, \neg P(x) \\ \neg\exists x\, P(x) &\Leftrightarrow \forall x\, \neg P(x) \end{aligned}$$

You may be able to see that these are true immediately. If not, here is an explanation of $\neg\forall x\, P(x) \Rightarrow \exists x\, \neg P(x)$ that should be convincing: If $\neg\forall x\, P(x)$, then $P(x)$ is not true for every x, which is to say that for some value a, $P(a)$ is not true. This means that $\neg P(a)$ is true. Since $\neg P(a)$ is true, it is certainly the case that there is some value of x that makes $\neg P(x)$ true, which is to say that $\exists x\, \neg P(x)$ is true. The other three implications may be explained in a similar way.

Here is another way to think of the quantifier versions of De Morgan's laws. The statement $\forall x\, P(x)$ is very much like a big conjunction. If the universe of discourse is the positive integers, for example, then it is equivalent to the statement that

$$P(1) \wedge P(2) \wedge P(3) \wedge \cdots$$

or, more concisely, we might write

$$\bigwedge_{x \in U} P(x),$$

using notation similar to "sigma notation" for sums. Of course, this is not really a "statement" in our official mathematical logic, because we don't allow infinitely long formulas. In the same way, $\exists x\, P(x)$ can be thought of as

$$\bigvee_{x \in U} P(x).$$

Now the first quantifier law can be written

$$\neg \bigwedge_{x \in U} P(x) \Leftrightarrow \bigvee_{x \in U} (\neg P(x)),$$

which looks very much like the law

$$\neg(P \wedge Q) \Leftrightarrow (\neg P \vee \neg Q),$$

but with an infinite conjunction and disjunction. Note that we can also rewrite De Morgan's laws for $\wedge$ and $\vee$ as

$$\neg \bigwedge_{i=1}^{2} (P_i(x)) \Leftrightarrow \bigvee_{i=1}^{2} (\neg P_i(x))$$
$$\neg \bigvee_{i=1}^{2} (P_i(x)) \Leftrightarrow \bigwedge_{i=1}^{2} (\neg P_i(x)).$$

This is more cumbersome, but it reflects the close relationship with the quantifier forms of De Morgan's laws.

Finally, general understanding is usually aided by specific examples: Suppose the universe is the set of cars. If $P(x)$ is "x has four wheel drive," then the denial of "every car has four wheel drive" is "there exists a car which does not have four wheel drive." This is an example of the first law. If $P(x)$ is "x has three wheels," then the denial of "there is a car with three wheels" is "every car does not have three wheels." This fits the pattern of the second law. In a more mathematical vein, a denial of the sentence "for every x, x^2 is positive" is "there is an x such that x^2 fails to be positive." A denial of "there is an x such that $x^2 = -1$" is "for every x, $x^2 \neq -1$."

It is easy to confuse the denial of a sentence with something stronger. If the universe is the set of all people, the denial of the sentence "All people are tall" is not the sentence "No people are tall." This might be called the **opposite** of the original sentence—it says more than simply "'All people are tall' is untrue." The correct denial of this sentence is "there is someone who is not tall," which is a considerably weaker statement. In symbols, the denial of $\forall x\ P(x)$ is $\exists x\ \neg P(x)$, whereas the opposite is $\forall x\ \neg P(x)$. ("Denial" is an "official" term in wide use; "opposite," as used here, is not widely used.)

De Morgan's laws can be used to simplify negations of the "some" form and the "all" form; the negations themselves turn out to have the same forms, but "reversed," that is, the negation of an "all" form is a "some" form, and vice versa. Suppose $P(x)$ and $Q(x)$ are formulas. We then have

$$\neg\forall x\ (P(x) \Rightarrow Q(x)) \ \Leftrightarrow\ \exists x\ (P(x) \wedge \neg Q(x))$$

$$\neg\exists x\ (P(x) \wedge Q(x)) \ \Leftrightarrow\ \forall x\ (P(x) \Rightarrow \neg Q(x))$$

The denial of the sentence "all lawn mowers run on gasoline" is the sentence "some lawn mower does not run on gasoline" (not "no lawn mowers run on gasoline," the opposite).

We verify the first statement and leave the second for an exercise:

$$\neg\forall x\ (P(x) \Rightarrow Q(x)) \Leftrightarrow \exists x\ \neg(P(x) \Rightarrow Q(x)) \Leftrightarrow \exists x\ (P(x) \wedge \neg Q(x))$$

A formula is usually simpler if $\neg$ does not appear in front of any compound expression, that is, it appears only in front of simple statements such as $P(x)$. The following is an example of simplifying the denial of a formula using De Morgan's laws:

$$\begin{aligned}\neg\forall x\ (P(x) \vee \neg Q(x)) &\Leftrightarrow \exists x\ \neg(P(x) \vee \neg Q(x)) \\ &\Leftrightarrow \exists x\ (\neg P(x) \wedge \neg\neg Q(x)) \\ &\Leftrightarrow \exists x\ (\neg P(x) \wedge Q(x))\end{aligned}$$

Denials of formulas are extremely useful. In a later section we will see that the techniques called proof by contradiction and proof by contrapositive use them extensively. Denials can also be a helpful study device. When you read a theorem or a definition in mathematics it is frequently helpful to form the denial of that sentence to see what it means for the condition to fail. The more ways you think about a concept in mathematics, the clearer it should become.

Augustus De Morgan. (y–1871; De Morgan himself noted that he was x years old in the year x^2.) De Morgan's father died when he was ten, after which he was raised by his mother, a devout member of the Church of England, who wanted him to be a minister. Far from becoming a minister, De Morgan developed a pronounced antipathy toward the Church, which would profoundly influence the course of his career.

De Morgan's interest in and talent for mathematics did not become evident until he was fourteen, but already at sixteen he entered Trinity College at Cambridge, where he studied algebra under George Peacock and logic under William Whewell. He was also an excellent flute player, and became prominent in musical clubs at Cambridge.

On graduation, De Morgan was unable to secure a position at Oxford or Cambridge, as he refused to sign the required religious test (a test not abolished until 1875). Instead, at the age of 22, he became Professor of Mathematics at London University, a new institution founded on the principle of religious neutrality.

De Morgan wrote prolifically about algebra and logic. Peacock and Gregory had already focused attention on the fundamental importance to algebra of symbol manipulation; that is, they established that the fundamental operations of algebra need not depend on the interpretation of the variables. De Morgan went one (big) step further: he recognized that the operations ($+$, $-$, etc.) also need have no fixed meaning (though he made an

exception for equality). Despite this view, De Morgan does seem to have thought that the only appropriate interpretations for algebra were familiar numerical domains, primarily the real and complex numbers. Indeed, he thought that the complex numbers formed the most general possible algebra, because he could not bring himself to abandon the familiar algebraic properties of the real and complex numbers, like commutativity.

One of De Morgan's most widely known books was *A Budget of Paradoxes*. He used the word 'paradox' to mean anything outside the accepted wisdom of a subject. Though this need not be interpreted pejoratively, his examples were in fact of the 'mathematical crank' variety—mathematically naive people who insisted that they could trisect the angle or square the circle, for example.

De Morgan's son George was himself a distinguished mathematician. With a friend, he founded the London Mathematical Society and served as its first secretary; De Morgan was the first president.

In 1866, De Morgan resigned his position to protest an appointment that was made on religious grounds, which De Morgan thought abused the principle of religious neutrality on which London University was founded. Two years later his son George died, and shortly thereafter a daughter died. His own death perhaps hastened by these events, De Morgan died in 1871 of 'nervous prostration.'

The information here is taken from *Lectures on Ten British Mathematicians*, by Alexander Macfarlane, New York: John Wiley & Sons, 1916.

Exercises 1.3.

1. Verify these tautologies using truth tables.

$$\neg(P \vee Q) \Leftrightarrow (\neg P \wedge \neg Q)$$
$$\neg(P \wedge Q) \Leftrightarrow (\neg P \vee \neg Q)$$

2. Suppose $R(x)$ is the statement "x is a rectangle," and $S(x)$ is the statement "x is a square." Write the following symbolically and decide which pairs of statements are denials of each other:

a) All rectangles are squares.

b) Some rectangles are squares.

c) Some squares are not rectangles.

d) No squares are rectangles.

e) No rectangles are squares.

f) All squares are rectangles.

g) Some squares are rectangles.

h) Some rectangles are not squares.

3. Write symbolically the following denials of definitions concerning a function f:

a) f is not increasing.

b) f is not decreasing.

c) f is not constant.

d) f does not have a root.

4. Simplify the following expressions:

a) $\neg\forall x > 0\ (x^2 > x)$

b) $\neg\exists x \in [0,1]\ (x^2 + x < 0)$

c) $\neg\forall x\ \forall y\ (xy = y^2 \Rightarrow x = y)$

d) $\neg\exists x\ \exists y\ (x > y \wedge y > x)$

5. Verify the statement:

$$\neg\exists x\ (P(x) \wedge Q(x)) \ \Leftrightarrow\ \forall x\ (P(x) \Rightarrow \neg Q(x))$$

6. Observe that

$$P \vee Q \Leftrightarrow \neg\neg(P \vee Q) \Leftrightarrow \neg(\neg P \wedge \neg Q)$$

so $\vee$ can be expressed in terms of $\wedge$ and $\neg$.

a) Show how to express $\Rightarrow$ in terms of $\wedge$ and $\neg$.

b) Show how to express $\wedge$ in terms of $\neg$ and $\vee$.

c) Show how to express $\vee$ in terms of $\neg$ and $\Rightarrow$.

7. Express the universal quantifier $\forall$ in terms of $\exists$ and $\neg$. Express $\exists$ in terms of $\forall$ and $\neg$.

8. Compute the year y of De Morgan's birth.

1.4 Mixed Quantifiers

In many of the most interesting mathematical formulas some variables are universally quantified and others are existentially quantified. You should be very careful when this is the case; in particular, the order of the quantifiers is extremely important. The universe in the following examples is the set of real numbers, except as noted.

EXAMPLE 1.4.1 Compare these two valid sentences:

$$\exists x\,\forall y\,(x + y = y), \qquad \forall y\,\exists x\,(x + y = 0).$$

In the first we require that x be a *fixed* value that satisfies the equation regardless of the value of y; clearly $x = 0$ will do. In the second formula, however, x *depends on* y; if $y = 3$, $x = -3$, if $y = 0$, $x = 0$. □

EXAMPLE 1.4.2 Compare these two sentences:

$$\forall x\,\exists y\,(y^3 = x), \qquad \exists y\,\forall x\ (xy^3 = -x).$$

The first is valid because given any x we can set y equal to the cube root of x. So as x varies, y also varies, that is, y depends upon x. The second is valid because there is a single fixed value $y = -1$ which makes the equation $xy^3 = -x$ valid, regardless of the value of x. □

EXAMPLE 1.4.3 Compare these two sentences:

$$\forall x\, \exists y\, (x < y), \qquad \exists y\, \forall x\, (x < y).$$

The first sentence is true and states that given any number there is a strictly larger number, that is, there is no largest number. The second sentence is false; it says that there is a single number that is strictly larger than all real numbers. □

In general, if you compare $\exists y\, \forall x\, P(x, y)$ with $\forall x\, \exists y\, P(x, y)$ it is clear that the first statement implies the second. If there is a fixed value y_0 which makes $P(x, y)$ true for all x, then no matter what x we are given, we can find a y (the fixed value y_0) which makes $P(x, y)$ true. So the first is a **stronger** statement. As in example 1.4.3, it is usually the case that this implication cannot be reversed.

We turn to some examples using more variables:

EXAMPLE 1.4.4 The sentence "between any two numbers is another number" can be written

$$\forall x\, \forall y\, \exists z\, ((x < y) \Rightarrow (x < z < y)).$$

Observe that z depends in an essential way on both variables "to its left," namely, x and y. The sentence is true if $U = \mathbb{R}$, but neither of these is true for $\mathbb{R}$: $\forall x\, \exists z\, \forall y\, ((x < y) \Rightarrow (x < z < y))$, $\exists z\, \forall x\, \forall y\, ((x < y) \Rightarrow (x < z < y))$. □

EXAMPLE 1.4.5 Suppose the universe of discourse is the integers. These are valid sentences:

$$\forall x\, \exists y\, \exists z\, (x = 7y + 5z), \qquad \forall x\, \exists y\, \forall z\, (z > x \Leftrightarrow z \geq y).$$

Consider the first sentence. If we know the value of x, we can choose $y = -2x$ and $z = 3x$, so $7y + 5z = -14x + 15x = x$. Notice that y and z depend on x in an essential way. Turning to the second, if we know x, we can choose y to be the next integer, $x + 1$. Any z is strictly larger than x if and only if it is at least as large as y. □

We often need to form denials of sentences with mixed quantifiers. These are handled with De Morgan's laws, just as in section 1.3.

EXAMPLE 1.4.6 The sentence $\exists x\, \forall y\, (x + y \neq 1)$ is false because its denial, $\forall x\, \exists y\, (x + y = 1)$, is valid. (For any number x, let $y = 1 - x$.) □

EXAMPLE 1.4.7 The sentence $\forall y\, \exists x\, (x^2 = y)$ is false because the denial of the sentence, $\exists y\, \forall x\, (x^2 \neq y)$, is valid. (Let $y = -1$.) □

EXAMPLE 1.4.8 If the universe is the integers, the sentence $\forall x\,\exists y\,\exists z\,(x = 4y + 6z)$ is false. Its denial is $\exists x\,\forall y\,\forall z\,(x \neq 4y + 6z)$. To see that this is valid, suppose x is any odd number. For any values of y and z, $4y + 6z$ is even, so it does not equal x. □

Exercises 1.4.

1. Using the real numbers as the universe of discourse, describe why the following are valid:
- **a)** $\exists x\,\forall y\,(xy = x^2)$
- **b)** $\forall x\,\exists y\,(x^2 + 6xy + 9y^2 = 0)$
- **c)** $\exists y\,\forall x\,(x + y > xy)$
- **d)** $\forall y\,\exists x\,(y - x = xy^2 + 1)$

2. Using the integers as the universe of discourse, describe why the following are valid:
- **a)** $\forall x\,\exists y\,\forall z\,(z < x \Leftrightarrow z \leq y)$
- **b)** $\forall x\,\exists y\,\exists z\,(x = 8y + 3z)$
- **c)** $\exists x\,\forall y\,\forall z\,(x = yz \Rightarrow y = -z)$
- **d)** $\exists x > 1\,\forall y\,\exists z\,((y = xz) \vee (y = xz + 1))$

3. Form the denials of the following and simplify using De Morgan's laws.
- **a)** $\forall x\,\exists y\,((x + y = 1) \wedge (xy \neq 0))$
- **b)** $\exists y\,\forall x\,((x^2 \neq y) \vee (x = y + 1))$
- **c)** $\forall x\,\exists y\,\forall z\,(x^2 + y = 0 \Rightarrow x < z^2)$
- **d)** $\exists x\,\forall y\,\exists z\,(((x + y = z) \wedge (x < y)) \Rightarrow ((x > z) \vee (y^2 = z)))$

4. Explain why the two sentences at the end of example 1.4.4 are false.

5. Why is the following valid (where U is the reals)?

$$\forall \epsilon{>}0\,\exists N{>}0\,\forall n\,(N < n \Rightarrow 1/n < \epsilon)$$

6. Using quantifiers, define what it means for $f\colon \mathbb{R} \to \mathbb{R}$ to be **periodic** (e.g., $\sin(x)$ is periodic). What does it mean for f to fail to be periodic? ($\mathbb{R}$ denotes the real numbers.)

7. Recall the famous quote by Abraham Lincoln:

> It is true that you may fool all the people some of the time; you can even fool some of the people all the time; but you can't fool all of the people all the time.

Let $F(x, y, z)$ mean "x can fool y at time z." Write Lincoln's statement symbolically. Identify some ambiguities in Lincoln's statement.

1.5 Logic and Sets

Like logic, the subject of **sets** is rich and interesting for its own sake. We will need only a few facts about sets and techniques for dealing with them, which we set out in this section and the next. We will return to sets as an object of study in chapters 4 and 5.

A set is a collection of objects; any one of the objects in a set is called a **member** or an **element** of the set. If a is an element of a set A we write $a \in A$.

Some sets occur so frequently that there are standard names and symbols for them. We denote the real numbers by $\mathbb{R}$, the rational numbers (that is, the fractions) by $\mathbb{Q}$, the integers by $\mathbb{Z}$ and the natural numbers (that is, the positive integers) by $\mathbb{N}$.

There is a natural relationship between sets and logic. If A is a set, then $P(x) =$"$x \in A$" is a formula. It is true for elements of A and false for elements outside of A. Conversely, if we are given a formula $Q(x)$, we can form the **truth set** consisting of all x that make $Q(x)$ true. This is usually written $\{x : Q(x)\}$ or $\{x \mid Q(x)\}$.

EXAMPLE 1.5.1 If the universe is $\mathbb{Z}$, then $\{x : x > 0\}$ is the set of positive integers and $\{x : \exists n\,(x = 2n)\}$ is the set of even integers. □

If there are a finite number of elements in a set, or if the elements can be arranged in a sequence, we often indicate the set simply by listing its elements.

EXAMPLE 1.5.2 $\{1, 2, 3\}$ and $\{1, 3, 5, 7, 9, \ldots\}$ are sets of integers. The second is presumably the set of all positive odd numbers, but of course there are an infinite number of other possibilities. In all but the most obvious cases, it is usually wise to describe the set ("the set of positive odd numbers, $\{1, 3, 5, 7, 9, \ldots\}$") or give a formula for the terms ("$\{1, 3, 5, 7, 9, \ldots, 2i + 1, \ldots\}$"). □

EXAMPLE 1.5.3 We indicate the **empty set** by $\emptyset$, that is, $\emptyset = \{\}$ is the set without any elements. Note well that $\emptyset \neq \{\emptyset\}$: the first contains nothing, the second contains a single element, namely the empty set. □

The logical operations $\neg, \wedge, \vee$ translate into the theory of sets in a natural way using truth sets. If A is a set, define

$$A^c = \{x : x \notin A\},$$

called the **complement** of A. If B is a second set, define

$$A \cap B = \{x : x \in A \wedge x \in B\},$$

called the **intersection** of A and B, and

$$A \cup B = \{x : x \in A \vee x \in B\},$$

called the **union** of A and B.

EXAMPLE 1.5.4 Suppose $U = \{1, 2, 3, \ldots, 10\}$, $A = \{1, 3, 4, 5, 7\}$, $B = \{1, 2, 4, 7, 8, 9\}$; then $A^c = \{2, 6, 8, 9, 10\}$, $A \cap B = \{1, 4, 7\}$ and $A \cup B = \{1, 2, 3, 4, 5, 7, 8, 9\}$. Note that the complement of a set depends on the universe U, while the union and intersection of two sets do not. □

We often wish to compare two sets. We say that A is a **subset** of B if

$$\forall x\ (x \in A \Rightarrow x \in B),$$

and write $A \subseteq B$. This is not only a definition but a technique of proof. If we wish to show $A \subseteq B$ we may start with an arbitrary element x of A and prove that it must be in B. We say the sets A and B are **equal** if and only if $A \subseteq B$ and $B \subseteq A$, that is,

$$\forall x\ (x \in A \Leftrightarrow x \in B).$$

So to show two sets are equal one must verify that a biconditional is satisfied, which often needs to be done in two parts, that is, the easiest way to show that $A = B$ often is to show that $A \subseteq B$ and $B \subseteq A$. If $A \subseteq B$ and $A \neq B$, we say A is a **proper** subset of B and write $A \subset B$.

EXAMPLE 1.5.5 $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. □

Finally, we say that A and B are **disjoint** if $A \cap B = \emptyset$.

In section 1.1 we learned that logical operations are related by many tautologies, the study of which is called Boolean Algebra. These tautologies can be interpreted as statements about sets; here are some particularly useful examples.

THEOREM 1.5.6 Suppose A, B and C are sets. Then

a) $A \cap B \subseteq A$,

b) $A \subseteq A \cup B$,

c) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,

d) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

e) $(A \cap B)^c = A^c \cup B^c$,

f) $(A \cup B)^c = A^c \cap B^c$,

g) $A \subseteq B$ iff $B^c \subseteq A^c$.

Proof. Suppose $P(x)$ ="$x \in A$", $Q(x)$ ="$x \in B$", $R(x)$ ="$x \in C$". To prove (a), suppose that $a \in A \cap B$. Then by definition, $P(a) \wedge Q(a)$ is true. Since $P(x) \wedge Q(x) \Rightarrow P(x)$ is a tautology, $P(a)$ is true, or $a \in A$. As noted above, this proves that $A \cap B \subseteq A$. Similarly, (c) follows since $P(x) \wedge (Q(x) \vee R(x)) \Leftrightarrow (P(x) \wedge Q(x)) \vee (P(x) \wedge R(x))$ is a tautology. All the other statements follow in the same manner. ■

As in the case of logic, (e) and (f) are called De Morgan's laws. Theorem 1.5.6 certainly is not an exhaustive list of set identities, it merely illustrates a few of the more important ones.

If $a, b \in U$ we can form the **ordered pair** (a, b). The fundamental property of ordered pairs is that $(a_1, b_1) = (a_2, b_2)$ if and only if $a_1 = a_2$ and $b_1 = b_2$. If A and B are sets, the set

$$A \times B = \{(a, b) : a \in A \wedge b \in B\}$$

is called the **Cartesian product** of A and B.

EXAMPLE 1.5.7 If $A = \{r, s, t\}$, $B = \{\$, \%\}$, then

$$A \times B = \{(r, \$), (r, \%), (s, \$), (s, \%), (t, \$), (t, \%)\}.$$

□

EXAMPLE 1.5.8 $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ is the plane. $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$ is 3-dimensional space. □

René Descartes. Descartes (1596–1650) was perhaps the most able mathematician of his time (though he may have to share top billing with Pierre de Fermat, a busy lawyer who did mathematics on the side for fun). Despite his ability and his impact on mathematics, Descartes was really a scientist and philosopher at heart. He made one great contribution to mathematics, *La géométrie*, and then concentrated his energies elsewhere.

La géométrie did not even appear on its own, but as an appendix to his most famous work, *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences* ("Discourse on the method of reasoning well and seeking truth in the sciences"). Descartes is remembered as the father of coordinate or analytic geometry, but his uses of the method were much closer in spirit to the great Greek geometers of antiquity than to modern usage. That is, his interest really lay in geometry; he viewed the introduction of algebra as a powerful tool for solving geometrical problems. Confirming his view that geometry is central, he went to some lengths to show how algebraic operations (for example, finding roots of quadratic equations) could be interpreted geometrically.

In contrast to modern practice, Descartes had no interest in graphing an arbitrary relation in two variables—in the whole of *La géométrie*, he did not plot any new curve from its equation. Further, ordered pairs do not play any role in the work; rectangular coordinates play no special role (Descartes used oblique coordinates freely—that is, his axes were not constrained to meet at a right angle); familiar formulas for distance, slope, angle between lines, and so on, make no appearance; and negative coordinates, especially negative abscissas, are little used and poorly understood. Ironically, then, there is little about the modern notion of Cartesian coordinates that Descartes would recognize.

Despite all these differences in emphasis and approach, Descartes' work ultimately made a great contribution to the theory of functions. The Cartesian product may be misnamed, but Descartes surely deserves the tribute.

Exercises 1.5.

1. For the given universe U and the given sets A and B, find A^c, $A \cap B$ and $A \cup B$.

a) $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $A = \{1, 3, 5, 8\}$, $B = \{2, 3, 5, 6\}$

b) $U = \mathbb{R}$, $A = (-\infty, 2]$, $B = (-1, \infty)$

c) $U = \mathbb{Z}$, $A = \{n : n \text{ is even}\}$, $B = \{n : n \text{ is odd}\}$

d) $U = \mathbb{Q}$, $A = \emptyset$, $B = \{q : q > 0\}$

e) $U = \mathbb{N}$, $A = \mathbb{N}$, $B = \{n : n \text{ is even}\}$

f) $U = \mathbb{R}$, $A = (-\infty, 0]$, $B = [-2, 3)$

g) $U = \mathbb{N}$, $A = \{n : n \leq 6\}$, $B = \{1, 2, 4, 5, 7, 8\}$

h) $U = \mathbb{R} \times \mathbb{R}$, $A = \{(x, y) : x^2 + y^2 \leq 1\}$, $B = \{(x, y) : x \geq 0, y \geq 0\}$.

2. Prove the parts of Theorem 1.5.6 not proved in the text.

3. Suppose U is some universe of discourse.

a) What is $\{x : x = x\}$?

b) What is $\{x : x \neq x\}$?

4. Prove carefully from the definition of "$\subseteq$" that for any set A, $\emptyset \subseteq A$.

5. **a)** If $A = \{1, 2, 3, 4\}$ and $B = \{x, y\}$, what is $A \times B$?

b) If A has m elements and B has n elements, how many elements are in $A \times B$?

c) Describe $A \times \emptyset$. Justify your answer.

d) What name do we give the set $(0, \infty) \times (0, \infty) \subset \mathbb{R}^2$?

e) What kind of geometric figure is $[1, 2] \times [1, 2] \subset \mathbb{R}^2$?

6. If A and B are sets, show $A \subseteq B$ iff $A \cap B^c = \emptyset$ iff $A^c \cup B = U$. (What are the corresponding logical statements?)

7. Suppose A, B, C and D are sets.

a) Show $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.

b) Does (a) hold with $\cap$ replaced by $\cup$?

8. Suppose we say a set S is **normal** if $S \notin S$. (You probably have encountered only normal sets, e.g., the *set* of reals is not a particular real.) Consider $N = \{S : S \text{ is a normal set}\}$. Is N a normal set? (This is called Russell's Paradox. Examples like this helped make set theory a mathematical subject in its own right. Although the concept of a set at first seems straightforward, even trivial, it emphatically is not.)

1.6 Families of Sets

Suppose I is a set, called the **index set**, and with each $i \in I$ we associate a set A_i. We call $\{A_i : i \in I\}$ an **indexed family of sets**. Sometimes this is denoted by $\{A_i\}_{i \in I}$.

EXAMPLE 1.6.1 Suppose I is the days of the year, and for each $i \in I$, A_i is the set of people whose birthday is i, so, for example, Beethoven $\in A_{(\text{December } 16)}$. □

EXAMPLE 1.6.2 Suppose I is the integers and for each $i \in I$, A_i is the set of multiples of i, that is, $A_i = \{x \in \mathbb{Z} : i|x\}$ (recall that $i|x$ means that x is a multiple of i; it is read "i divides x"). □

Given an indexed family $\{A_i : i \in I\}$ we can define the intersection and union of the sets A_i using the universal and existential quantifiers:

$$\bigcap_{i\in I} A_i = \{x : \forall i \in I\,(x \in A_i)\}$$
$$\bigcup_{i\in I} A_i = \{x : \exists i \in I\,(x \in A_i)\}.$$

EXAMPLE 1.6.3 If $\{A_i : i \in I\}$ is the indexed family of example 1.6.1, then $\bigcap_{i\in I} A_i$ is the empty set and $\bigcup_{i\in I} A_i$ is the set of all people. If $\{A_i : i \in I\}$ is the indexed family of example 1.6.2, then $\bigcap_{i\in I} A_i$ is $\{0\}$ and $\bigcup_{i\in I} A_i$ is the set of all integers. □

Since the intersection and union of an indexed family are essentially "translations" of the universal and existential quantifiers, it should not be too surprising that there are De Morgan's laws that apply to these unions and intersections.

THEOREM 1.6.4 If $\{A_i : i \in I\}$ is an indexed family of sets then

a) $(\bigcap_{i\in I} A_i)^c = \bigcup_{i\in I} A_i^c$,

b) $(\bigcup_{i\in I} A_i)^c = \bigcap_{i\in I} A_i^c$.

Proof. We'll do (a): $x \in (\bigcap_{i\in I} A_i)^c$ iff $\neg(x \in \bigcap_{i\in I} A_i)$ iff $\neg\forall i \in I\,(x \in A_i)$ iff $\exists i \in I\,(x \notin A_i)$ iff $\exists i \in I\,(x \in A_i^c)$ iff $x \in \bigcup_{i\in I} A_i^c$. ■

You may be puzzled by the inclusion of this theorem: is it not simply part of theorem 1.5.6? No: theorem 1.5.6 (parts (e) and (f)) concerns the intersection or union of two sets only. This can be extended easily to any intersection or union of a finite number of sets, though even this modest extension does require separate proof. The real problem is with intersections or unions of an infinite number of sets. Though in this case the extension to infinite operations has an easy proof, it is not always the case that what is true for a finite number of operations is true for an infinite number of operations, and even when true, the proof in the infinite case may be more difficult.

The relationships in the following theorem are simple but useful; they illustrate the dual nature of the union and intersection of families of sets.

THEOREM 1.6.5 If $\{A_i : i \in I\}$ is an indexed family of sets and B is any set, then

a) $\bigcap_{i\in I} A_i \subseteq A_j$, for each $j \in I$,

b) $A_j \subseteq \bigcup_{i\in I} A_i$, for each $j \in I$.

c) if $B \subseteq A_i$, for all $i \in I$, then $B \subseteq \bigcap_{i\in I} A_i$,

d) if $A_i \subseteq B$, for all $i \in I$, then $\bigcup_{i\in I} A_i \subseteq B$.

Proof. Part (a) is a case of **specialization**, that is, if $x \in \bigcap_{i\in I} A_i$, then $x \in A_i$ for all $i \in I$, in particular, when $i = j$. Part (d) also is easy—if $x \in \bigcup_{i\in I} A_i$, then for some $i \in I$, $x \in A_i \subseteq B$, so $x \in B$. Parts (b) and (c) are left as exercises. ■

An indexed family $\{A_i : i \in I\}$ is **pair-wise disjoint** if $A_i \cap A_j = \emptyset$ whenever i and j are distinct elements of I. The indexed family of example 1.6.1 is pair-wise disjoint, but the one in example 1.6.2 is not. If S is a set then an indexed family $\{A_i : i \in I\}$ of subsets of S is a **partition of** S if it is pair-wise disjoint and $S = \bigcup_{i\in I} A_i$. Partitions appear frequently in mathematics.

EXAMPLE 1.6.6 Let $I = \{e, o\}$, A_e be the set of even integers and A_o be the set of odd integers. Then $\{A_i : i \in I\}$ is a partition of $S = \mathbb{Z}$. □

EXAMPLE 1.6.7 Let $I = \mathbb{R}$, $S = \mathbb{R}^2$, and for each $i \in I$, let $A_i = \{(x, i) : x \in \mathbb{R}\}$. Each A_i is a horizontal line and the indexed family partitions the plane. □

Sometimes we want to discuss a collection of sets (that is, a set of sets) even though there is no natural index present. In this case we can use the collection itself as the index.

EXAMPLE 1.6.8 If $\mathcal{S} = \{\{1, 3, 4\}, \{2, 3, 4, 6\}, \{3, 4, 5, 7\}\}$, then $\bigcap_{A\in\mathcal{S}} A = \{3, 4\}$ and $\bigcup_{A\in\mathcal{S}} A = \{1, 2, 3, 4, 5, 6, 7\}$. □

An especially useful collection of sets is the **power set of a set**: If X is any set, the power set of X is $\mathcal{P}(X) = \{A : A \subseteq X\}$.

EXAMPLE 1.6.9 If $X = \{1, 2\}$, then $\mathcal{P}(X) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$. □

EXAMPLE 1.6.10 $\mathcal{P}(\emptyset) = \{\emptyset\}$, that is, the power set of the empty set is non-empty. □

Exercises 1.6.

1. Let $I = \{1, 2, 3\}$, $A_1 = \{1, 3, 4, 6, 7\}$ $A_2 = \{1, 4, 5, 7, 8, 9\}$, $A_3 = \{2, 4, 7, 10\}$. Find $\bigcap_{i\in I} A_i$ and $\bigcup_{i\in I} A_i$.

2. Suppose $I = [0, 1] \subseteq \mathbb{R}$ and for each $i \in I$, let $A_i = (i - 1, i + 1) \subseteq \mathbb{R}$. Find $\bigcap_{i\in I} A_i$ and $\bigcup_{i\in I} A_i$.

3. Prove parts (b) and (c) of theorem 1.6.5.

4. Suppose U is the universe of discourse and the index set $I = \emptyset$. What should we mean by $\bigcap_{i\in I} A_i$ and $\bigcup_{i\in I} A_i$? Show that Theorem 1.6.4 still holds using your definitions.

5. Suppose $\{A_i\}_{i\in I}$ is a partition of a set S. If $T \subseteq S$, show that $\{A_i \cap T\}_{i\in I}$ is a partition of T.

6. A collection of sets, $\mathcal{S}$, is **totally ordered** if for every $A, B \in \mathcal{S}$, either $A \subseteq B$ or $B \subseteq A$. If $\mathcal{S}$ is totally ordered, show that $\mathcal{S}^c = \{A^c : A \in \mathcal{S}\}$ is totally ordered.

7. Suppose $\mathcal{S}$ is a collection of sets and B is some other set. Show that if B is disjoint from every $A \in \mathcal{S}$ then B is disjoint from $\bigcup_{A\in\mathcal{S}} A$.

2

Proofs

Proof may be what best distinguishes mathematics from other disciplines, even the sciences, which are logical, rigorous and to a greater or lesser degree (depending on the discipline) based on mathematics. By using rigorous, logically correct reasoning, we aim to prove mathematical theorems—that is, to demonstrate that something is true beyond all doubt.

It is impossible to give a formula or algorithm for proving any and all mathematical statements, yet certain approaches or strategies appear over and over in successful proofs, so studying proof itself is worthwhile. Of course, even if the subject is proof itself, we need to prove *something*, so in this chapter we begin our study of **number theory**, that is, the properties of the integers (often, but not always, the non-negative integers).

The idea of proof is central to all branches of mathematics; we concentrate on proofs involving the integers for two reasons. First, it is a very good subject in which to learn to write proofs. The proofs in number theory are typically very clean and clear; there is little in the way of abstraction to cloud one's understanding of the essential points of an argument. Secondly, the integers have a central position in mathematics and are used extensively in other fields such as computer science. Although the great twentieth century mathematician G. H. Hardy boasted that he did number theory because there was no chance that it could be construed as applied mathematics, it has in fact become enormously useful and important in the study of computation and particularly in cryptography. We also find number theory intrinsically interesting, one of the most beautiful subjects in modern mathematics, and all the more interesting because of its roots in antiquity. Unless otherwise specified, then, the universe of discourse is the set of integers, $\mathbb{Z}$.

2.1 Direct Proofs

A **proof** is a sequence of statements. These statements come in two forms: **givens** and **deductions**. The following are the most important types of "givens."

Hypotheses: Usually the theorem we are trying to prove is of the form

$$P_1 \wedge \ldots \wedge P_n \Rightarrow Q.$$

The Ps are the hypotheses of the theorem. We can *assume* that the hypotheses are true, because if one of the P_i is false, then the implication is true.

Known results: In addition to any stated hypotheses, it is always valid in a proof to write down a theorem that has already been established, or an unstated hypothesis (which is usually understood from context).

Definitions: If a term is defined by some formula it is always legitimate in a proof to replace the term by the formula or the formula by the term.

We turn now to the most important ways a statement can appear as a consequence of (or deduction from) other statements:

Tautology: If P is a statement in a proof and Q is logically equivalent to P, we can then write down Q.

Modus Ponens: If the formula P has occurred in a proof and $P \Rightarrow Q$ is a theorem or an earlier statement in the proof, we can write down the formula Q. Modus ponens is used frequently, though sometimes in a disguised form; for example, most algebraic manipulations are examples of modus ponens.

Specialization: If we know "$\forall x\, P(x)$," then we can write down "$P(x_0)$" whenever x_0 is a particular value. Similarly, if "$P(x_0)$" has appeared in a proof, it is valid to continue with "$\exists x\, P(x)$". Frequently, choosing a useful special case of a general proposition is the key step in an argument.

When you read or write a proof you should always be very clear *exactly* why each statement is valid. You should always be able to identify how it follows from earlier statements.

A **direct proof** is a sequence of statements which are either givens or deductions from previous statements, and whose last statement is the conclusion to be proved.

Variables: The proper use of variables in an argument is critical. Their improper use results in unclear and even incorrect arguments. Every variable in a proof has a quantifier associated with it, so there are two types of variables: those that are universally quantified and those that are existentially quantified. We may fail to mention explicitly how a variable is quantified when this information is clear from the context, but every variable has an associated quantifier.

A universally quantified variable is introduced when trying to prove a statement of the form $\forall x(P(x) \Rightarrow Q(x))$. The language typically employed is "Suppose x satisfies $P(x)$," "Assume $P(x)$," or "Let $P(x)$."

When we introduce an existentially quantified variable, it is usually defined in terms of other things that have been introduced previously in the argument. In other words, it depends on the previously mentioned quantities in the proof.

DEFINITION 2.1.1 We say the integer n is **even** if there is an integer k such that $n = 2k$. We say n is **odd** if there is a k such that $n = 2k - 1$. □

EXAMPLE 2.1.2 If n is even, so is n^2.

Proof. Assume n is an even number (n is a universally quantified variable which appears in the statement we are trying to prove). Because n is even, $n = 2k$ for some k (k is existentially quantified, defined in terms of n, which appears previously). Now $n^2 = 4k^2 = 2(2k^2)$ (these algebraic manipulations are examples of modus ponens). Let $j = 2k^2$ (j is existentially quantified, defined in terms of k); then $n^2 = 2j$, so n is even (by definition). ■

EXAMPLE 2.1.3 The sum of two odd numbers is even.

Proof. Assume m and n are odd numbers (introducing two universally quantified variables to stand for the quantities mentioned in the statement). Because m and n are odd there are integers j and k such that $m = 2j - 1$ and $n = 2k - 1$ (introducing existentially quantified variables, defined in terms of quantities already mentioned). Now $m + n = (2j - 1) + (2k - 1) = 2(j + k - 1)$ (modus ponens). Let $i = j + k - 1$ (existentially quantified); then $m + n = 2i$ is even (by definition). ■

Exercises 2.1.

In 1-4, write proofs for the given statements, inserting parenthetic remarks to explain the rationale behind each step (as in the examples).

1. The sum of two even numbers is even.
2. The sum of an even number and an odd number is odd.
3. The product of two odd numbers is odd.
4. The product of an even number and any other number is even.
5. Suppose in the definitions of even and odd the universe of discourse is assumed to be the real numbers, $\mathbb{R}$, instead of the integers. What happens? What if the universe is the natural numbers, $\mathbb{N}$? (When we say the universe is U we mean that both n and k in the definition are from U.)
6. Prove that x is odd if and only if $|x|$ is odd.
7. If x and y are integers and $x^2 + y^2$ is even, prove that $x + y$ is even.

2.2 Divisibility

If $n \neq 0$ and a are integers, we say that n **divides** a (and write $n|a$) if there exists an m such that $a = nm$. When $n|a$ we also say n is a **divisor** of a and a is a **multiple** of n.

A word of caution: The symbol $n|a$ is *not* a fraction, but a formula. It means that there is a relationship between the two numbers which is either true or false (2 and 6 have this relationship, 2 and 7 do not). While we are studying number theory we will have no occasion to mention the rational numbers—we will, in fact, avoid them. There are several reasons for this. One is practical: a given fraction has more than one representation (e.g., $4/12 = 5/15$). It is also possible that a number that doesn't look like an integer is, in fact, an integer (e.g., $45/15$). These ambiguities can be a real source of confusion. A second reason is theoretical: the integers can be used to define the other number systems (more on this in chapter 5), so the integers should be studied as a self-contained subject before dealing with these other systems. A third reason is aesthetic: number theory is the study of the *integers*; it is somehow more elegant and satisfying to provide proofs that use only number theoretic results and techniques. (We do not mean to overstate this. Mathematics is a single discipline, and some of the most beautiful and elegant proofs bring apparently unrelated parts of mathematics together to solve a problem. These surprising connections between different parts of mathematics enhance the whole mathematical enterprise.)

The following theorem will be very useful, despite its simplicity.

THEOREM 2.2.1 If $n|a$ and $n|b$ then $n|ax + by$ for any $x, y \in \mathbb{Z}$, so in particular $n|(a + b), n|(a - b)$ and $n|ax$.

Proof. Suppose $a = ni$, $b = nj$. Then $ax + by = nix + njy = n(ix + jy)$ which shows that $n|ax + by$. The second statement contains particular instances of the first, where in the first case $x = 1$, $y = 1$, in the second case $x = 1$, $y = -1$ and in the third case $y = 0$. ■

COROLLARY 2.2.2 If $n|a$ and $a|b$, then $n|b$.

Proof. Since $a|b$, there is an x such that $b = ax$. Now the result follows from 2.2.1. ■

An integer $p > 0$ is called **prime** if it has exactly two positive divisors, namely, 1 and p. If $a > 0$ has more than two positive divisors, we say it is **composite**. It is important, but easy to forget, that 1 is not prime (neither is it composite). A prime has exactly two positive divisors, but 1 has only one (1 itself). Observe that if $a > 1$ is composite, then $a = nm$ where $n, m > 1$ (just let n be any positive divisor other than 1 and a).

There are many theorems about primes that are amazing (some are amazingly hard to prove). There are also many questions involving primes which, though they are easy to state, have resisted all attempts at proof. A simple one has to do with so-called **twin**

primes, pairs of primes of the form p and $p+2$ (e.g. 5 & 7, 11 & 13). No one knows whether there are an infinite number of such pairs, though they occur as far out as anyone has checked (by computer). There also are some arguments that make it appear likely that the number of twin primes is infinite.

Exercises 2.2.

1. For the given n, a, show $n|a$ by finding an m with $a = nm$.
- **a)** $4|20$
- **b)** $5|-25$
- **c)** $-3|9$
- **d)** $-9|-27$
- **e)** $1|23$
- **f)** $-1|17$
- **g)** $-5|0$
- **h)** $75|75$

2. Prove, directly from the definition of '|', that for any integer $x \neq 0$, $x|0$, $1|x$ and $x|x$.

3. Find all integers n such that $n|(2n+3)$.

4. Prove that if $m|a$ and $n|b$, then $mn|ab$.

5. Show that if $m \neq 0$, then $nm|am$ iff $n|a$.

6. Prove that if $a|b$, then $|a| \Big| |b|$.

7. If n is an integer, let (n) be the set of all multiples of n, i.e., $(n) = \{a : n|a\}$.
- **a)** If a,b are in (n) and x and y are any integers, prove $ax + by$ is in (n).
- **b)** If $(n) \subseteq (m)$, prove $m|n$.
- **c)** If $m|n$, prove $(n) \subseteq (m)$.

2.3 Existence proofs

Many interesting and important theorems have the form $\exists x\, P(x)$, that is, that there exists an object x satisfying some formula P. In such **existence proofs**, try to be as specific as possible. The most satisfying and useful existence proofs often give a concrete example, or describe explicitly how to produce the object x.

EXAMPLE 2.3.1 To prove the statement, *there is a prime number p such that $p+2$ and $p+6$ are also prime numbers*, note that $p = 5$ works because $5+2 = 7$ and $5+6 = 11$ are also primes. □

In this example, 5 is not the only number that works (e.g., 11 works as well). In fact, it is a famous unsolved problem whether there are infinitely many primes that work. This

would be a more interesting theorem, but the point remains: when doing an existence proof, be as concrete as possible.

EXAMPLE 2.3.2 Suppose U is a universe which is appropriate for a calculus class. To prove the statement, *there is a function f such that $f' = f$*, note that $f(x) = e^x$ works (as does any constant multiple of e^x). □

A slight variation on the existence proof is the *counter-example*. Suppose you look at a sentence of the form $\forall x P(x)$ and you come to the conclusion that it is false and you wish to prove this. What you wish to prove, then, is $\neg\forall x P(x)$, which by De Morgan's law is equivalent to $\exists x \neg P(x)$. A specific x satisfying $\neg P(x)$ is called a counter-example to the assertion that $\forall x P(x)$.

EXAMPLE 2.3.3 To disprove the sentence *for every integer x, if $6x$ is even then x is even*, we write it symbolically as

$$\forall x\,(6x \text{ is even} \Rightarrow x \text{ is even}).$$

To disprove this, note that $x = 3$ is a counter-example, because $6 \cdot 3$ is even but 3 is odd. It may help, especially in more complicated examples, to form the denial explicitly:

$$\exists x\,(6x \text{ is even} \wedge x \text{ is not even}).$$

Now it is easy to see that $x = 3$ makes the formula in parentheses true. □

EXAMPLE 2.3.4 Suppose U is the universe of example 2.3.2. To disprove the sentence *for every function f, if f is continuous at 0 then it is differentiable at 0*, note that $f(x) = |x|$ is a counter-example. □

Once again, the most satisfying way to prove something false is to come up with a specific counter-example, though sometimes this is difficult or impossible. Note well that it is never sufficient simply to find an error in the proof of some sentence to conclude that it is false—it is easy to come up with erroneous proofs of correct facts. If you have trouble proving a statement of the form $\forall x\, P(x)$, try looking at some particular cases of the result. You may find a counter-example, or you may get a hint about why the statement really is true.

There are occasions when it is impossible, or very difficult, to find a specific example. An existence proof sometimes can be constructed by indirect means, or by using other existence results.

EXAMPLE 2.3.5 Using a universe as in example 2.3.2, show *there is a solution for the equation $x^3 + 3x - 2 = 0$ in the interval $[0, 1]$*. Let $f(x) = x^3 + 3x - 2$; then since $f(0) = -2$

and $f(1) = 2$, the Intermediate Value Theorem says that there is an x in $[0, 1]$ for which $f(x) = 0$. □

If you consult a good calculus text, you should find that the Mean Value Theorem (which is an existence result), is proved by referring to Rolle's Theorem (another existence result), which is proved by referring to the Maximum Value Theorem (yet a third existence result, sometimes called the Extreme Value Theorem), which is proved "indirectly," without ever exhibiting the object that is claimed to exist. At no point are we given a formula for the quantity we seek, and the result is perhaps not as satisfying as we would like. In general, then, try to be specific when doing an existence proof, but if you cannot, it may still be possible to construct an example using some other existence result or another technique of proof.

Trying to prove a statement of the form $\forall x \exists y P(x, y)$ is rather like trying to do many existence arguments at the same time. For any value x we would like to construct or describe a y that makes $P(x, y)$ true.

EXAMPLE 2.3.6 *There is no largest integer.* We want to prove $\forall n \exists m(n < m)$. Argue as follows: Suppose $n \in \mathbb{Z}$ is given. Let $m = n + 1$. Then $n < n + 1 = m$. □

EXAMPLE 2.3.7 $\forall x \exists y(2y = x^2 + x)$. Let x be an arbitrary integer. If x is even, x^2 is even and if x is odd, x^2 is odd. Since the sum of two even numbers or two odd numbers is even, $x^2 + x$ is even. Therefore, $\exists y(2y = x^2 + x)$, by the definition of even, and in fact $y = (x^2 + x)/2$. □

EXAMPLE 2.3.8 *There are arbitrarily long gaps in the sequence of prime numbers.* In other words, we want to prove that for every positive integer n there is a positive integer m such that $m + 1$, $m + 2$, ..., $m + n$ are all composite. For any n, let $m = (n + 1)! + 1 = (n + 1)n(n - 1) \cdots 1 + 1$. If $1 \leq k \leq n$, then $m + k = (n + 1)! + (k + 1)$. Since both $(n + 1)!$ and $k + 1$ are divisible by $k + 1 > 1$, $m + k$ is composite. □

Exercises 2.3.

The universe of discourse is shown in parentheses.

1. ($\mathbb{N}$) Show that there is a prime number p such that $p + 4$ and $p + 6$ are also prime numbers.
2. ($\mathbb{N}$) Show that there are prime numbers p and q such that $p + q = 128$. (This is a case of the famous **Goldbach Conjecture**, which says that every even integer $n \geq 4$ can be written as the sum of two primes. It seems highly probable from work with computers that the Goldbach Conjecture is true, but no one has discovered a proof.)
3. ($\mathbb{Z}$) Show that every odd integer is the sum of two consecutive integers.
4. ($\mathbb{Z}$) Show that every odd integer is the difference between two consecutive perfect squares.
5. ($\mathbb{N}$) Disprove the following: If $12|x^2$, then $12|x$.

6. ($\mathbb{N}$) Disprove the following: If $x|ab$, then $x|a$ or $x|b$.

7. ($\mathbb{N}$) Disprove the following: If $n^2|m^3$, then $n|m$.

8. ($\mathbb{R}$) Show there is an $x \in [0, \pi/2]$ such that $\cos x = x$.

2.4 Induction

Perhaps you have seen the method of **proof by induction** before. Stated in the abstract, or exhibited in a simple example, it is easy to understand and seems hardly worth much attention. Yet induction is an extraordinarily powerful and subtle method of proof. We will use a version of induction that probably is different than what you have seen before.

DEFINITION 2.4.1 Induction Axiom Suppose that $P(n)$ is a formula and m and $k \geq 0$ are fixed integers. Suppose further that

1. $P(m), P(m+1), \ldots, P(m+k)$ are all true, and

2. for every $n > m+k$, the implication $P(m), \ldots, P(n-1) \Rightarrow P(n)$ is valid.

Then $P(n)$ is true for all $n \geq m$. □

When $k = 0$ this is often called **complete induction**. You may be more familiar with the simplest form of induction, where $m = 1$, $k = 0$, and the implication in (2) is replaced with $P(n-1) \Rightarrow P(n)$. Clearly, though, these forms of induction express the same basic idea: If some statement is true about "small" integers, and if knowing that the statement is true up to some integer $n-1$ always allows you to prove that it is true for n, then the statement must be true for all n (more precisely, all n larger than the "small" integers you started with). In effect, induction is a way to describe an infinite number of proofs simultaneously: how to prove $P(2)$ knowing $P(1)$, how to prove $P(3)$ knowing $P(1)$ and $P(2)$, how to prove $P(4)$ knowing $P(1)$ and $P(2)$ and $P(3)$, and so on. This only works if all the proofs are essentially the same—it is conceivable that $P(n)$ is true for all n, yet the proofs are much different for different values of n. In such an unhappy circumstance, induction will not be much help.

Proofs by induction always include verification of (1) and (2). Usually the first is called the **base case** or the **basis** of the induction, and the second is called the **induction step**. To prove the induction step, assume that $P(m), \ldots, P(n-1)$ are all true and try to prove $P(n)$. The statements $P(m), \ldots, P(n-1)$ are called the **induction hypothesis**.

EXAMPLE 2.4.2 For every $n \geq 1$,

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

Proof. We apply induction with $m = 1$ and $k = 0$ to this formula. Clearly

$$\sum_{i=1}^{1} i = 1 = \frac{1 \cdot 2}{2},$$

proving the base case. Now, assuming that $n > 1$ and $P(1), \ldots, P(n-1)$ are true, then in particular $P(n-1)$ is true, that is,

$$\sum_{i=1}^{n-1} i = \frac{(n-1)n}{2},$$

so

$$\sum_{i=1}^{n} i = n + \sum_{i=1}^{n-1} i = n + \frac{(n-1)n}{2} = \frac{n(n+1)}{2}.$$

■

This example employs the simplest kind of induction, since $k = 0$ and we only needed the one case, $P(n-1)$, of the induction hypothesis. The power of the more general form is that it allows us to assume the validity of the proposition for *all* values less than n, which is very useful in many proofs.

EXAMPLE 2.4.3 Every non-negative integer is either even or odd.

Proof. What we wish to establish is that every non-negative n can be expressed as $2q+r$, where $r = 0$ (the even case) or $r = -1$ (the odd case). We apply induction with $m = 0$ and $k = 1$. Clearly $0 = 2 \cdot 0$ and $1 = 2 \cdot 1 - 1$. Now assume that $n \geq 2$ and that the result is true for $0, 1, \ldots, n-1$. Then $n - 2 \geq 0$ so there are integers q' and r such that $n - 2 = 2q' + r$, with $r = 0$ or -1. Set $q = q' + 1$; then $2q + r = 2(q' + 1) + r = 2q' + r + 2 = (n-2) + 2 = n$, as desired. ■

The following result is part of the **Fundamental Theorem of Arithmetic**; we will see the rest of the proof in section 3.5.

THEOREM 2.4.4 Every integer $n \geq 2$ can be factored into a product of primes.

Proof. We use induction with $m = 2$. Clearly $n = 2$ can be factored into a product of primes (it is already prime), so the base case is true. For the induction step, assume $n > 2$ and that $2, \ldots, n-1$ can each be factored into primes; we need to show that this implies that n can also be factored into primes. We divide this into two cases: if n is a prime, then it already is factored as desired. If n is not a prime, then it factors as $a \cdot b$, where $2 \leq a, b < n$. Since we have assumed that all numbers between 2 and $n-1$ factor into

primes, there are primes $p_1, \ldots, p_i$ and $q_1, \ldots, q_j$ such that $a = p_1 \cdots p_i$ and $b = q_1 \cdots q_j$, and so

$$n = a \cdot b = p_1 \cdots p_i \cdot q_1 \cdots q_j,$$

that is, n can be written as a product of primes. ■

This proof is a very natural one, because it mimics the way most people would in fact factor a number: Faced with such a problem, you probably would try to factor the number in any way at all, writing $n = ab$. If it does not factor it must be prime and you are done. You then start over with these smaller numbers a and b and try to factor them. Each time you perform this operation, the factors get smaller, so you are assured that eventually the process must stop. It is possible in such circumstances to say something like this by way of proof: "Look, just keep doing this, and eventually it stops, and then the result is true." This is essentially proof by induction, but more informally stated. In simple cases such a proof may be acceptable, but in more difficult circumstances, it will be harder to see that such a process works. It really is best to practice doing proof by induction in the formal, "official" manner, so that you will be prepared to write and understand more difficult proofs.

COROLLARY 2.4.5 Every integer $n \geq 2$ is divisible by some prime.

Proof. Factor n into primes and grab any prime in the list. ■

Exercises 2.4.

1. Prove that

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

2. Prove that

$$1 + 4 + 9 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

3. Prove that any non-negative integer can be expressed as $3q + r$, where $0 \leq r < 3$.

4. Prove that any $n \geq 8$ can be expressed as $3x + 5y$ where $x \geq 0$ and $0 \leq y < 3$.

5. Suppose $a_0 = 1$, $a_1 = 2$ and for every $n > 1$, $a_n = 3a_{n-1} - 2a_{n-2}$. Find a simple formula for the value of a_n, and prove that it is correct.

6. Show that $2^{(2^n)} + 1$ is a prime for $n = 0, 1, 2, 3, 4$. Show that it is not prime when $n = 5$. (You may use computing devices.)

7. The Fibonacci numbers are defined by $F_0 = 0$, $F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$. Prove that $F_{m+n} = F_{m+1}F_n + F_mF_{n-1}$ for $n \geq 1$ and $m \geq 0$.

8. Use the previous problem to prove that $F_n | F_{kn}$.

9. Is $n^2 + n + 41$ prime for every integer $n \geq 0$? If so, prove it; if not, find the *smallest* n for which it is composite. (You may use computing devices.)

In the remaining problems, use the universe that is implied in the statement of the problem.

10. Assuming you know that the derivative of a constant is 0 and the derivative of x is 1, use induction and the product rule to prove $(x^n)' = nx^{n-1}$ for all $n \geq 0$.

11. A polynomial is **irreducible** if it cannot be factored into two polynomials of strictly smaller degree. Show that any polynomial is the product of irreducible polynomials. (Hint: let $P(n)$ be the formula "all polynomials of degree n can be factored into a product of irreducible polynomials".)

12. In a round robin tennis tournament, every player plays every other player once. Let's say that a "winner" is any player x such that for every other player y, either x beat y or else there is a player z such that x beat z and z beat y. In a round robin tournament with at least 2 players, show there is at least one winner.

13. A polygon in the plane is **convex** if the segment connecting any two vertices is contained entirely inside the polygon. Use induction to prove that the sum of the n angles of a convex polygon with n vertices is $(n-2)\pi$. You may assume that the sum of the angles in a triangle is π.

2.5 Uniqueness Arguments

Some of the most useful and interesting existence theorems are "existence and uniqueness proofs"—they say that there is one and only one object with a specified property. The symbol $\exists! x\ P(x)$ stands for "there exists a unique x satisfying $P(x)$," or "there is exactly one x such that $P(x)$," or any equivalent formulation.

EXAMPLE 2.5.1 (The universe is $\mathbb{R}$.) $\exists! x\,(x^2 + 1 = 2x)$: This is true since $x = 1$ is not only a solution, but the *only* solution. (Can you prove it?) □

We can, of course, combine this quantifier with others.

EXAMPLE 2.5.2 (The universe is $\mathbb{Z}$.) $\forall x\,\exists! y\,(x < y < x + 2)$: This is true since only $y = x + 1$ satisfies the inequalities. □

The quantifier $\exists!$ can be broken down into the "existence" part and the "uniqueness" part. In other words, $\exists x!\, P(x)$ says the same thing as

$$\big(\exists x\, P(x)\big) \wedge \big(\forall x\, \forall y\, (P(x) \wedge P(y) \Rightarrow x = y)\big)$$

The second part of this formula is the "uniqueness" part; it says that any two elements that satisfy P must, in fact, be the same. More often than not, we must prove existence and uniqueness separately; quite frequently, the uniqueness part is the easier of the two.

EXAMPLE 2.5.3 (The universe is $\mathbb{R}$.) *There is a unique function $f(x)$ such that $f'(x) = 2x$ and $f(0) = 3$.*

Proof. *Existence*: $f(x) = x^2 + 3$ works.

Uniqueness: If $f_0(x)$ and $f_1(x)$ both satisfy these conditions, then $f_0'(x) = 2x = f_1'(x)$, so they differ by a constant, i.e., there is a C such that $f_0(x) = f_1(x) + C$. Hence, $3 = f_0(0) = f_1(0) + C = 3 + C$. This gives $C = 0$ and so $f_0(x) = f_1(x)$. ■

Sometimes we can do both parts of an existence and uniqueness argument at the same time. This is usually accomplished by proving $\forall x\,(P(x) \Leftrightarrow x = x_0)$, where x_0 is some particular value.

EXAMPLE 2.5.4 *For every x there exists a unique y such that $(x+1)^2 - x^2 = 2y - 1$.*

Proof. We do this in two ways—first we divide up the argument:
Existence: Let $y = x + 1$; then $(x+1)^2 - x^2 = x^2 + 2x + 1 - x^2 = 2(x+1) - 1 = 2y - 1$.
Uniqueness: If y_0 and y_1 both satisfy the equation, then $2y_0 + 1 = (x+1)^2 - x^2 = 2y_1 + 1$, so $2y_0 + 1 = 2y_1 + 1$. Subtracting 1 and canceling 2 gives $y_0 = y_1$.
We can also combine existence and uniqueness:

Note that $(x+1)^2 - x^2 = x^2 + 2x + 1 - x^2 = 2(x+1) - 1$. Now $2y - 1 = 2(x+1) - 1$ if and only if $y = x + 1$, which proves the result.

(We're cheating a little here. In fact, proving a statement of the form $(P(x) \Leftrightarrow x = x_0)$ often requires two proofs, one for each direction of the "$\Leftrightarrow$". Sometimes, as in this case, the proof can be phrased so that the "if and only if" is clear without two distinct proofs. In general, however, an existence and uniqueness proof is likely to require two proofs, whichever way you choose to divide the work.) ■

Here is a familiar yet extraordinarily useful existence and uniqueness theorem, called the *Division Algorithm*. It says that if we divide one integer into another we end up with a unique quotient and remainder.

THEOREM 2.5.5 If a and b are integers and b is positive, then there are integers q and r such that $a = bq + r$ and $0 \le r < b$. Furthermore, these numbers (called the quotient and remainder) are unique.

Proof. We begin with the existence part of the argument. We assume $a \ge 0$ and leave the case $a < 0$ for an exercise. We use induction on a, with $m = 0$, and $k = b - 1$. (Recall the use of m and k in the statement of the Induction Axiom in section 2.4.)

If $0 \le a \le b - 1$, then $a = b \cdot 0 + a$; this establishes the basis of the induction. Now assume that $a \ge b$ and the result is true for $0, 1, \ldots, a-1$. Then $a - 1 \ge a - b \ge 0$, so there are numbers q' and $r \in \{0, 1, \ldots, b-1\}$ such that $a - b = bq' + r$. Let $q = q' + 1$; then

$$qb + r = (q' + 1)b + r = q'b + r + b = (a - b) + b = a,$$

as desired.

To show uniqueness, suppose

$$bq_1 + r_1 = a = bq_2 + r_2, \quad \text{where} \quad 0 \le r_1 \le r_2 < b.$$

Then

$$0 \le r_2 - r_1 < b - 0 = b,$$

because $r_2 < b$ and $r_1 \ge 0$. We also have

$$b(q_1 - q_2) = bq_1 - bq_2 = r_2 - r_1,$$

which gives

$$0 \le b(q_1 - q_2) < b.$$

Canceling the b's gives

$$0 \le q_1 - q_2 < 1,$$

i.e. $q_1 - q_2 = 0$, or $q_1 = q_2$. This, in turn, implies $0 = r_2 - r_1$, i.e., $r_1 = r_2$. Note that the uniqueness portion of the proof does not use the hypothesis that $a \ge 0$, so all that remains is to show existence for $a < 0$ (exercise 7). ■

You may object that this is an awful lot of work for such an "obvious" result. But this result almost certainly seems obvious because it is so familiar—you know by experience that you can always get a quotient and remainder. Yet pressed to explain how you know that no matter how you do it, you always get the *same* remainder, you would probably find yourself at something of a loss. As we set out to establish a body of true mathematical facts, it is important to have complete confidence that the foundation is solid. Many a convincing proof has turned out to be wrong; the first place to look for a mistake is *always* the line that says "now, it is obvious that . . . "

COROLLARY 2.5.6 Using the above notation, $b|a$ if and only if $r = 0$.

Proof. See exercise 8. ■

EXAMPLE 2.5.7 Suppose $a = 11$ and $b = 3$. You undoubtedly know how to find the quotient q: divide a by b and round down: $\lfloor 11/3 \rfloor = 3$ (the notation $\lfloor x \rfloor$ indicates the **floor** function, or the **greatest integer** function, that is, the "round down to the nearest integer" function). Now $bq = 9$ and $r = 2$, giving $11 = 3 \cdot 3 + 2$. This also works when $a < 0$, say $a = -11$ and $b = 3$: $\lfloor -11/3 \rfloor = \lfloor -3.66\ldots \rfloor = -4$, so $q = -4$, $bq = -12$, $r = 1$, and $-11 = 3 \cdot -4 + 1$. □

Exercises 2.5.

In 1-4, identify the existence part and the uniqueness part of your proof clearly.

1. There is a unique solution to $2x - 3 = 7$.

2. For every x there is a unique y such that $(x+1)^3 - x^3 = 3y + 1$.

In the next two exercises, assume the universe of discourse is appropriate for a calculus class.

3. There is a unique function f such that $f'(x) = \sin x$, $f(\pi/2) = 0$.

4. There is a unique function f such that $f'(x) = f(x)$ and $f(0) = 1$ (note $f(x) = e^x$ works. To show uniqueness differentiate $f(x)e^{-x}$).

5. For the following values of a and b, find q and r such that $0 \leq r < b$ and $a = qb + r$.

a) $a = 81$, $b = 6$

b) $a = 728$, $b = 7$

c) $a = -11$, $b = 8$

d) $a = -58$, $b = 9$

e) $a = 375$, $b = 1$

f) $a = 5$, $b = 11$

6. Find a positive integer a and integers q and q' so that

$$a = 5q + 1 = 7q' + 3.$$

7. Here we extend the proof of theorem 2.5.5 to the case of negative numbers. Suppose $-a$ is a negative number (so a is positive). We already know that there are unique q' and r such that $a = q'b + r'$ and $0 \leq r' < b$. To finish the proof, we need to find q and r such that $-a = qb + r$. Hint: Consider two cases, when $r' = 0$ and when $r' > 0$. These examples may help:

$$25 = 2 \cdot 9 + 7 \quad -25 = (-3) \cdot 9 + 2$$
$$49 = 5 \cdot 9 + 4 \quad -49 = (-6) \cdot 9 + 5$$

8. Provide the details of the proof of corollary 2.5.6.

2.6 Indirect Proof

Quite frequently you will find that it is difficult (or impossible) to prove something directly, but easier (at least possible) to prove it *indirectly*. The essence of the idea is simple: for example, suppose you want to know whether it is overcast or sunny, but you can't see the sky through your window. You usually can tell, indirectly, by the quality of light that you *can* see. Without formalizing the process, you make use of something like the following: If it is sunny I will be able to see areas of bright light and areas of shadow in the garden; I don't, so it must be (at least partially) overcast.

There are two methods of indirect proof: proof of the contrapositive and proof by contradiction. They are closely related, even interchangeable in some circumstances, though

proof by contradiction is more powerful. What unites them is that they both start by *assuming the denial of the conclusion.*

Proof of the Contrapositive

The contrapositive of the statement $P \Rightarrow Q$ is $\neg Q \Rightarrow \neg P$. For example, the contrapositive of "If it is Sunday, I go to church" is "If I am not going to church, it is not Sunday." Any sentence and its contrapositive are logically equivalent (theorem 1.1.3), but often it is easier and more natural to prove the contrapositive of a sentence.

EXAMPLE 2.6.1 *If $n > 0$ and $4^n - 1$ is prime, then n is odd:* Assume $n = 2k$ is even. Then

$$4^n - 1 = 4^{2k} - 1 = (4^k - 1)(4^k + 1).$$

Therefore, $4^n - 1$ factors (are both factors bigger than 1?) and hence is not prime. □

EXAMPLE 2.6.2 *If ab is even then either a or b is even:* Assume both a and b are odd. Since the product of odd numbers is odd, ab is odd. □

Proof by Contradiction:

To prove a sentence P by contradiction we assume $\neg P$ and derive a statement that is known to be false. Since mathematics is consistent (at least we hope so), this means P must be true.

In the case that the sentence we are trying to prove is of the form $P \Rightarrow Q$, we assume that P is true and Q is false (because $P \wedge \neg Q$ is the negation of $P \Rightarrow Q$), and try to derive a statement known to be false. Note that this statement need not be $\neg P$—this is the principal difference between proof by contradiction and proof of the contrapositive. In a proof of the contrapositive, we assume that Q is false and try to prove that P is false.

EXAMPLE 2.6.3 $\sqrt{3} \notin \mathbb{Q}$: Assume $\sqrt{3} = a/b$ for positive integers a and b with no common factors (i.e., a/b is in "lowest terms"). Then $a^2/b^2 = 3$, so $a^2 = 3b^2$. Now $3|3b^2$ so $3|a^2$. This implies that $3|a$, so $a = 3k$ for some k. Then $a^2 = (3k)^2 = 9k^2 = 3b^2$, or $3k^2 = b^2$. Now $3|b^2$, so $3|b$. Thus we've shown that 3 divides both a and b, but this contradicts the fact that a/b is in lowest terms. Hence, $\sqrt{3}$ cannot be written as a ratio of whole numbers. (We haven't proved that if $3|n^2$ then $3|n$, which we used twice. This can be proved in much the same way that we proved facts about even and odd numbers in section 2.1.) □

Proof by contradiction makes some people uneasy—it seems a little like magic, perhaps because throughout the proof we appear to be 'proving' false statements. A direct proof, or even a proof of the contrapositive, may seem more satisfying. Still, there seems to be

no way to avoid proof by contradiction. (Attempts to do so have led to the strange world of "constructive mathematics".)

The following simple but wonderful proof is at least as old as Euclid's book *The Elements*.

THEOREM 2.6.4 There are infinitely many primes.

Proof. Assume there are only finitely many primes $p_1, \ldots, p_k$. Let $n = p_1 \cdots p_k + 1$. Clearly $n \geq 2$, so by corollary 2.4.5, n is divisible by some prime, say p_i. Obviously, $p_i | (p_1 \cdots p_i \cdots p_k)$, so by theorem 2.2.1, $p_i | (n - p_1 \cdots p_k)$; Since $n - p_1 \cdots p_k = 1$, $p_i | 1$, a contradiction. ■

Note that this theorem does not give us a formula for constructing an infinite list of prime numbers. In particular, n itself is not necessarily prime, though of course it might be. To date no one has devised a prime-generating formula, though many have tried.

Are there any clues that might lead you to think that an indirect proof might be a good idea? The presence of *not*'s in the statement of a theorem we are trying to prove is often (but not always!) an indication that an indirect argument is worth trying. Theorem 2.6.4 says that a certain set is *not finite*. Example 2.6.2 has the form $P \Rightarrow (Q \vee R)$. By denying the conclusion, we get two 'solid facts' to work with: $\neg Q$ and $\neg R$. Working with composite numbers is often easier than working with primes, because the composite number can be factored; if by using indirect proof you can introduce a composite number it may help. Unfortunately, there are no hard and fast rules—deciding on what approach to use is a matter of experience and trial and error.

Euclid of Alexandria. Euclid, who flourished around 300 BC, is known to most high school students as the father of geometry. Surprisingly little is known of his life, not even his dates or birthplace. Shortly before 300 BC, Ptolemy I founded the great university at Alexandria, the first institution of its kind, and not unlike the universities of today. Euclid was recruited, probably from Athens, to head the mathematics department.

Euclid appears to have been primarily a teacher, not a great originator of new material. His *Elements*, unquestionably the most successful textbook of all time, often is thought to be an encyclopedia of all geometrical knowledge at the time. In fact, it is an elementary textbook covering geometry, arithmetic and algebra; Euclid himself knew and wrote about more advanced topics in mathematics. The perception that the *Elements* is only about geometry presumably is due to two facts: his name is most closely associated with geometry in modern elementary mathematics; and the mathematicians of antiquity, lacking

modern algebraic notation, did all arithmetic and algebra in the language of geometry—for example, numbers were not thought of in the abstract, but as the lengths of line segments, or measures of areas or volumes.

The *Elements* consists of thirteen books containing much that is still familiar to students: most of elementary geometry, of course, including the Pythagorean Theorem; the theorem on the number of primes and the *Fundamental Theorem of Arithmetic;* and the *Euclidean Algorithm*, which we will see in section 3.3.

Two famous stories are told about Euclid. It is said that Ptolemy asked him if geometry could be learned without reading the *Elements,* to which Euclid replied, "There is no royal road to geometry." (This story is also told about Menaechmus and Alexander the Great, which perhaps diminishes its credibility somewhat.) In response to a student who questioned the use of geometry, Euclid reportedly ordered that the student be given three pence, "since he must needs make gain of what he learns."

For more information, see *A History of Mathematics,* by Carl B. Boyer, New York: John Wiley and Sons, 1968; or *An Introduction to the History of Mathematics,* by Howard Eves, New York: Holt, Rinehart and Winston, 1976.

Exercises 2.6.

1. If $n > 0$ and $6^n - 1$ is prime, prove that n is odd.

2. If $a + b$ is odd, prove that a or b is odd.

3. Prove that $\sqrt{2}$ is not a rational number. Hint: use the results of section 2.1.

4. Prove that $\sqrt{8}$ is not a rational number.

5. If $a + b > 100$, prove that either $a > 50$ or $b > 50$.

6. Using $\mathbb{R}$ as the universe, prove that if a is a rational number and b is not a rational number, then $a + b$ is not a rational number.

7. An integer n is said to be **square-free** if it has no divisors that are perfect squares (other than 1). Show that any divisor of a square-free integer is square-free.

8. Show that for every integer $n > 2$ there is a prime between n and $n! = 1 \cdot 2 \cdots (n-1) \cdot n$. (Hint: look for prime factors of $n! - 1$.)

9. Prove that if $3|n^2$ then $3|n$.

3

Number Theory

3.1 Congruence

As with so many concepts we will see, **congruence** is simple, perhaps familiar to you, yet enormously useful and powerful in the study of number theory. If n is a positive integer, we say the integers a and b are **congruent** modulo n, and write $a \equiv b \pmod{n}$, if they have the same remainder on division by n. (By remainder, of course, we mean the unique number r defined by the Division Algorithm.) This notation, and much of the elementary theory of congruence, is due to the famous German mathematician, Carl Friedrich Gauss—certainly the outstanding mathematician of his time, and perhaps the greatest mathematician of all time.

EXAMPLE 3.1.1 $\{\ldots, -6, 1, 8, 15, \ldots\}$ are all congruent modulo 7 because their remainders on division by 7 equal 1. $\{\ldots, -4, 4, 12, 20, \ldots\}$ are all congruent modulo 8 since their remainders on division by 8 equal 4. □

Here is a wonderfully useful result.

LEMMA 3.1.2 $a \equiv b \pmod{n}$ if and only if $n|(a-b)$.

Proof. We break the proof into two parts:

(only if) If $a \equiv b \pmod{n}$, then there are integers q, q' and r, with $a = qn + r$ and $b = q'n + r$. So $a - b = (qn + r) - (q'n + r) = (q - q')n$, which means $n|a - b$.

(if) Suppose $n|a - b$, so there is an x with $a - b = xn$, that is, $a = b + xn$. Suppose r is the remainder on dividing n into b; we need to show that r is also the remainder on

dividing n into a. Since $b = qn + r$, we have $a = b + xn = qn + r + xn = (q + x)n + r$. Thus, when n is divided into a, the remainder is r as desired. ■

If the value of n is clear from the context, we often write simply $a \equiv b$. Congruence of integers shares many properties with equality; we list a few here.

THEOREM 3.1.3 Congruence modulo n satisfies the following:

1. $a \equiv a$ for any a;
2. $a \equiv b$ implies $b \equiv a$;
3. $a \equiv b$ and $b \equiv c$ implies $a \equiv c$;
4. $a \equiv 0$ iff $n|a$;
5. $a \equiv b$ and $c \equiv d$ implies $a + c \equiv b + d$;
6. $a \equiv b$ and $c \equiv d$ implies $a - c \equiv b - d$;
7. $a \equiv b$ and $c \equiv d$ implies $ac \equiv bd$;
8. $a \equiv b$ implies $a^j \equiv b^j$ for each integer $j \geq 1$.

Proof. Parts 1, 2, 3 and 4 are clear by the definition of congruence. (Aren't they? Check!) We'll prove parts 6 and 8, leaving parts 5 and 7 as exercises. Part 6: By hypothesis $n|a - b$ and $n|c - d$, so we have $n|(a - b) - (c - d)$. Rearranging the terms, this means $n|(a - c) - (b - d)$, so $a - c \equiv b - d$. Part 8: This follows from part 7, but it is easy to prove it directly: since $a \equiv b$, $n|a - b$. Therefore,

$$n|(a - b)(a^{j-1} + a^{j-2}b + \ldots + ab^{j-2} + b^{j-1}) = a^j - b^j,$$

so $a^j \equiv b^j$. Be sure you notice how often we have used lemma 3.1.2. ■

Parts 5–8 can be summarized by saying that in any expression involving $+, -, \cdot$ and positive integer exponents (that is, any "polynomial"), if individual terms are replaced by other terms that are congruent to them modulo n, the resulting expression is congruent to the original.

EXAMPLE 3.1.4 Any perfect square is of the form $4x$ or $4x + 1$, that is, if you divide 4 into a perfect square, the remainder is never 2 or 3: Suppose k^2 is some perfect square. Then k is congruent modulo 4 to exactly one of $0, 1, 2$ or 3, so k^2 is congruent to $0^2 = 0$, $1^2 = 1$, $2^2 \equiv 0$ or $3^2 \equiv 1$, so it is never congruent to 2 or 3. □

EXAMPLE 3.1.5 Find all integers x such that $3x-5$ is divisible by 11. Put in somewhat more familiar terms, we are trying to solve the congruence $3x \equiv 5 \pmod{11}$ for x, much as we might try to solve an equation for an unknown. Let's assume $3x \equiv 5$ and see what

that tells us about x. Since $4 \cdot 3 = 12 \equiv 1$,

$$3x \equiv 5 \Rightarrow 4 \cdot 3x \equiv 4 \cdot 5 \Rightarrow 12x \equiv 20 \Rightarrow x \equiv 9.$$

So if $3x \equiv 5$ then $x \equiv 9$, or $x \in \{\ldots, -13, -2, 9, 20, \ldots\}$. We also want to know that in fact all of these values *are* solutions, that is, if $x \equiv 9$ then $3x \equiv 5$. This is easy. (Right?) □

EXAMPLE 3.1.6 You are probably familiar with the old rule ("casting out nines") that an integer is divisible by 9 if and only if the sum of its digits is divisible by 9. Here is a proof. Suppose x is some positive integer and when we write it in decimal form it looks like $d_k d_{k-1} \ldots d_1 d_0$ (where each d_i is between 0 and 9). This means

$$x = d_k \cdot 10^k + d_{k-1} \cdot 10^{k-1} + \ldots + d_1 \cdot 10 + d_0.$$

Observe that $10 \equiv 1 \pmod 9$ and so $10^i \equiv 1^i = 1 \pmod 9$ for every i. This implies that

$$x \equiv d_k + d_{k-1} + \ldots + d_1 + d_0 \pmod 9.$$

This actually proves more than we need. It says that an integer and the sum of its digits are congruent modulo 9. In particular, one is congruent to 0 (that is, divisible by 9) if and only if the other is. □

Carl Friedrich Gauss. Gauss (1777–1855) was an infant prodigy and arguably the greatest mathematician of all time (if such rankings mean anything; certainly he would be in almost everyone's list of the top five mathematicians, as measured by talent, accomplishment and influence). Perhaps the most famous story about Gauss relates his triumph over busywork. As Carl Boyer tells the story: "One day, in order to keep the class occupied, the teacher had the students add up all the numbers from one to a hundred, with instructions that each should place his slate on a table as soon as he had completed the task. Almost immediately Carl placed his slate on the table, saying, 'There it is;' the teacher looked at him scornfully while the others worked diligently. When the instructor finally looked at the results, the slate of Gauss was the only one to have the correct answer, 5050, with no further calculation. The ten-year-old boy evidently had computed mentally the sum of the arithmetic progression $1 + 2 + \cdots + 100$, presumably through the formula $m(m + 1)/2$."

By the time Gauss was about 17, he had devised and justified the method of least squares, but had not decided whether to become a mathematician or a philologist. Just short of his nineteenth birthday, he chose mathematics, when he succeeded in constructing (under the ancient restriction to compass and straightedge) a seventeen-sided regular

polygon, the first polygon with a prime number of sides to be constructed in over 2000 years; previously, only the equilateral triangle and the regular pentagon had been constructed. Gauss later proved precisely which regular polygons can be constructed. (The answer is somewhat unsatisfying, however. He proved that the regular polygons that can be constructed have $2^m p_1 p_2 \cdots p_r$ sides, for any $m \geq 0$ and distinct **Fermat primes** p_i, that is, prime numbers having the form $2^{2^n} + 1$ for some n. Unfortunately, it is not known whether there are an infinite number of Fermat primes.)

Gauss published relatively little of his work, but from 1796 to 1814 kept a small diary, just nineteen pages long and containing 146 brief statements. This diary remained unknown until 1898. It establishes in large part the breadth of his genius and his priority in many discoveries. Quoting Boyer again: "The unpublished memoranda of Gauss hung like a sword of Damocles over mathematics of the first half of the nineteenth century. When an important new development was announced by others, it frequently turned out that Gauss had had the idea earlier, but had permitted it to go unpublished."

The range of Gauss's contributions is truly stunning, including some deep and still standard results such as the *Quadratic Reciprocity Theorem* and the *Fundamental Theorem of Algebra*. He devoted much of his later life to astronomy and statistics, and made significant contributions in many other fields as well. His name is attached to many mathematical objects, methods and theorems; students of physics may know him best as the namesake of the standard unit of magnetic intensity, the **gauss**.

The information here is taken from *A History of Mathematics*, by Carl Boyer, New York: John Wiley & Sons, 1968.

Exercises 3.1.

1. For the given values of n and a, find the number $b \in \{0, 1, \ldots, n-1\}$ for which $a \equiv b \pmod{n}$.

a) $n = 7$, $a = 30$

b) $n = 9$, $a = 69$

c) $n = 2$, $a = 123{,}472{,}461$

d) $n = 6$, $a = -60$

e) $n = 11$, $a = -63$

f) $n = 17$, $a = -38$

2. If $a = nq + r$, it is not necessarily the case that r is the remainder on dividing a by n; for example, $20 = 6 \cdot 2 + 8$, but 8 is certainly not the remainder when we divide 20 by 6. In lemma 3.1.2 we showed that $a = (q + x)n + r$ and concluded from this that the remainder on dividing a by n is r. Explain why this conclusion is justified.

3. Prove parts (5) and (7) in theorem 3.1.3. (For part 7, you might want to prove $ac \equiv bc \equiv bd$.)

4. Prove part (8) from part (7) in theorem 3.1.3, by induction.

5. What digits can appear in the 1's place of a perfect square?
6. Prove that $x \equiv 9 \pmod{11}$ iff $3x \equiv 5 \pmod{11}$.
7. Find all x such that $7x + 3$ is divisible by 9.
8. Suppose n and m are positive integers. Show that
$$ma \equiv mb \pmod{mn} \quad \Leftrightarrow \quad a \equiv b \pmod{n}.$$
(See exercise 5 of section 2.2.)
9. State and prove a result similar to example 3.1.6 regarding divisibility by 11.
10. Prove that for any integer x, $x^3 - x$ is divisible by 6.
11. Find a rule, similar to example 3.1.6, that determines when a three-digit number is divisible by 7, and prove that it works.
12. Find the remainder when 111111110888888895 is divided by 9.

3.2 $\mathbb{Z}_n$

We saw in theorem 3.1.3 that when we do arithmetic modulo some number n, the answer doesn't depend on which numbers we compute with, only that they are the same modulo n. For example, to compute $16 \cdot 30 \pmod{11}$, we can just as well compute $5 \cdot 8 \pmod{11}$, since $16 \equiv 5$ and $30 \equiv 8$. This suggests that we can go further, devising some universe in which there really is no difference between 16 and 5 (assuming that we want to work modulo 11).

Throughout this section, unless otherwise specified, assume all equivalences are modulo n, for some fixed but unspecified n.

DEFINITION 3.2.1 For every integer a, let $[a]$ have the property that $[a] = [a']$ if and only if $a \equiv a'$. □

Note that this is a very peculiar definition: we give no hint as to what $[a]$ *is*—we only specify one aspect of its behavior. This turns out not to matter much, but we will eventually see what $[a]$ "really" is.

Recall that if r is the remainder on dividing n into a, then $a \equiv r$, or, in our new language, $[a] = [r]$. This means that every $[a]$ is equal to some $[r]$ for $0 \leq r < n$; this motivates the next definition.

DEFINITION 3.2.2 Let $\mathbb{Z}_n = \{[0], [1], [2], \ldots, [n-1]\}$. □

That is, by the preceding remark, $\mathbb{Z}_n$ consists of all possible $[a]$. This is a new universe in which we can investigate "arithmetic".

EXAMPLE 3.2.3 $\mathbb{Z}_4 = \{[0], [1], [2], [3]\}$. We could write $\mathbb{Z}_4 = \{[-80], [25], [102], [-13]\}$ instead, but only to make a point—not in practice. □

EXAMPLE 3.2.4 In $\mathbb{Z}_5$, $[1] = [6] = [-4]$, $[3] = [8] = [-2]$. □

Now we're ready to see if we can indeed do arithmetic in this new universe $\mathbb{Z}_n$. We start with the simplest operations, namely, addition, subtraction and multiplication.

DEFINITION 3.2.5 If $[a], [b] \in \mathbb{Z}_n$, let $[a] + [b] = [a + b]$, $[a] - [b] = [a - b]$ and $[a] \cdot [b] = [ab]$. □

Most mathematicians would agree that these definitions are natural, even inevitable. You might try to think of other ways that these simple operations might be defined on $\mathbb{Z}_n$.

EXAMPLE 3.2.6 Here are the addition and multiplication tables for $\mathbb{Z}_4$.

$+$	$[0]$	$[1]$	$[2]$	$[3]$
$[0]$	$[0]$	$[1]$	$[2]$	$[3]$
$[1]$	$[1]$	$[2]$	$[3]$	$[0]$
$[2]$	$[2]$	$[3]$	$[0]$	$[1]$
$[3]$	$[3]$	$[0]$	$[1]$	$[2]$

$\times$	$[0]$	$[1]$	$[2]$	$[3]$
$[0]$	$[0]$	$[0]$	$[0]$	$[0]$
$[1]$	$[0]$	$[1]$	$[2]$	$[3]$
$[2]$	$[0]$	$[2]$	$[0]$	$[2]$
$[3]$	$[0]$	$[3]$	$[2]$	$[1]$

□

Unfortunately, though we have characterized the definitions of addition, subtraction and multiplication as "natural," the situation is not as straightforward as it may first appear. The definition $[a] + [b] = [a + b]$ depends on the manipulation of specific integers a and b, but we know that there are other integers c and d with $[a] = [c]$ and $[b] = [d]$. What if we compute $[c + d]$? We had better get the same result as $[a + b]$ or the definition of addition doesn't make sense: $[a] + [b]$ would be different than $[c] + [d]$, but they must be the same. Fortunately, theorem 3.1.3 comes to the rescue, and the two quantities $[a + b]$ and $[c + d]$ are the same. Here's why: since $[a] = [c]$ and $[b] = [d]$, a and c are congruent modulo n, as are b and d. Therefore their sums $a + b$ and $c + d$ are congruent which means that $[a + b] = [c + d]$. Subtraction and multiplication can be justified in the same way. What we have shown is that the definitions of addition, subtraction and multiplication are "well-defined."

Many of the familiar algebraic properties of integers carry over to $\mathbb{Z}_n$; here are a few of the most familiar.

THEOREM 3.2.7 In $\mathbb{Z}_n$,

a) $[a] + [b] = [b] + [a]$,

b) $[a] + ([b] + [c]) = ([a] + [b]) + [c]$,

c) $[a] \cdot [b] = [b] \cdot [a]$,

d) $[a] \cdot ([b] \cdot [c]) = ([a] \cdot [b]) \cdot [c]$,

e) $[a] \cdot ([b] + [c]) = [a] \cdot [b] + [a] \cdot [c]$.

f) $[0] + [a] = [a]$,

g) $[0] \cdot [a] = [0]$,

h) $[1] \cdot [a] = [a]$.

Proof. We prove two parts and leave the rest as exercises.

Part (a) follows since $[a] + [b] = [a + b] = [b + a] = [b] + [a]$; in other words, we just reduce it to the corresponding statement for regular addition. Similarly (f) follows since $[0] + [a] = [0 + a] = [a]$. ■

Parts (a) and (c) are commutative laws, (b) and (d) are associative laws and (e) says that multiplication distributes over addition. Parts (f), (g) and (h) show that [0] and [1] act in $\mathbb{Z}_n$ in much the same way that 0 and 1 act in $\mathbb{Z}$. Though many properties of the integers are shared by $\mathbb{Z}_n$, there are exceptions; here is one.

EXAMPLE 3.2.8 In $\mathbb{Z}$, if $ab = 0$ then either a or b must be 0, but in $\mathbb{Z}_n$ this need not be the case. For example, in $\mathbb{Z}_{12}$, $[3] \cdot [4] = [12] = [0]$, but $[3] \neq [0]$ and $[4] \neq [0]$. □

We do not yet know what $[a]$ is, but it certainly is not an integer, so $\mathbb{Z}_n$ is not a subset of $\mathbb{Z}$. Remember this well; it sometimes is tempting to confuse $\mathbb{Z}_n = \{[0], [1], [2], \ldots, [n-1]\}$ with $\{0, 1, 2, \ldots, n-1\} \subset \mathbb{Z}$. The brackets make all the difference in the world: in $\mathbb{Z}_5$, $[2] = [7]$, but of course $2 \neq 7$.

Exercises 3.2.

1. Construct addition and multiplication tables for

a) $\mathbb{Z}_2$

b) $\mathbb{Z}_6$.

2. Prove the remaining parts of Theorem 3.2.7.

3. If $[a]$ and $[b]$ are in $\mathbb{Z}_n$, prove that there is a unique $[x] \in \mathbb{Z}_n$ such that $[a] + [x] = [b]$.

4. Use the table from exercise 1(b) to verify the following statements:

a) There is a unique $[x] \in \mathbb{Z}_6$ such that $[5] \cdot [x] = [2]$

b) There is no $[x] \in \mathbb{Z}_6$ such that $[3] \cdot [x] = [4]$.

c) There is an $[x] \in \mathbb{Z}_6$ such that $[4] \cdot [x] = [2]$, but it is not unique.

5. Find all the elements $[x]$ of $\mathbb{Z}_{15}$ such that $[x] = [p]$ for some prime number p (p need not be less than 15).

6. Suppose you add together all the elements of $\mathbb{Z}_n$. What is the result?

7. In $\mathbb{Z}_{12}$, find all of the elements $[x]$ such that $[x]^n = [0]$ for some positive integer n.

3.3 The Euclidean Algorithm

Suppose a and b are integers, not both zero. The **greatest common divisor** (gcd, for short) of a and b, written (a, b) or $\gcd(a, b)$, is the largest positive integer that divides both a and b. We will be concerned almost exclusively with the case where a and b are non-negative, but the theory goes through with essentially no change in case a or b is negative. The notation (a, b) might be somewhat confusing, since it is also used to denote ordered pairs and open intervals. The meaning is usually clear from the context. We begin with some simple observations:

LEMMA 3.3.1 Suppose a and b are not both zero.

a) $(a, b) = (b, a)$,

b) if $a > 0$ and $a|b$ then $(a, b) = a$,

c) if $a \equiv c \pmod{b}$, then $(a, b) = (c, b)$.

Proof. Part (a) is clear, since a common divisor of a and b is a common divisor of b and a. For part (b), note that if $a|b$, then a is a common divisor of a and b. Clearly a is the largest divisor of a, so we are done. Finally, if $a \equiv c \pmod{b}$, then $b|a - c$, so there is a y such that $a - c = by$, i.e., $c = a - by$. If d divides both a and b, then it also divides $a - by$. Therefore any common divisor of a and b is also a common divisor of c and b. Similarly, if d divides both c and b, then it also divides $c + by = a$, so any common divisor of c and b is a common divisor of a and b. This shows that the common divisors of a and b are exactly the common divisors of c and b, so, in particular, they have the same greatest common divisor. ■

It perhaps is surprising to find out that this lemma is all that is necessary to compute a gcd, and moreover, to compute it very efficiently. This remarkable fact is known as the **Euclidean Algorithm**. As the name implies, the Euclidean Algorithm was known to Euclid, and appears in *The Elements*; see section 2.6. As we will see, the Euclidean Algorithm is an important theoretical tool as well as a practical algorithm. Here is how it works:

To compute (a, b), divide the larger number (say a) by the smaller number, so $a = bq_1 + r_1$ and $r_1 < b$. By 3.3.1(c), $(a, b) = (b, r_1)$. Now $b = r_1q_2 + r_2$, $r_2 < r_1$, and $(b, r_1) = (r_1, r_2)$; then $r_1 = r_2q_3 + r_3$, $r_3 < r_2$, and $(r_1, r_2) = (r_2, r_3)$, and so on. Since $r_1 > r_2 > r_3 \ldots$, eventually some $r_k = 0$ and $(a, b) = (r_{k-1}, r_k) = (r_{k-1}, 0) = r_{k-1}$; in other words, (a, b) is the last non-zero remainder we compute. Note that $(a, 0) = a$, by 3.3.1(b).

EXAMPLE 3.3.2

$$\begin{aligned}(198, 168) &= (168, 30)\\ &= (30, 18)\\ &= (18, 12)\\ &= (12, 6)\\ &= (6, 0) = 6.\end{aligned}$$

□

If you have done some computer programming, you should see just how easy it is to implement this algorithm in any reasonable programming language. Since it is a very fast algorithm it plays an important role in many applications.

With a little extra bookkeeping, we can use the Euclidean Algorithm to show that $\gcd(a, b)$ is actually a **linear combination** of a and b.

EXAMPLE 3.3.3 Again taking $a = 198$ and $b = 168$:

$$\begin{aligned}30 &= 198 - 168 = a - b,\\ 18 &= 168 - 5 \cdot 30 = b - 5(a - b) = -5a + 6b,\\ 12 &= 30 - 18 = (a - b) - (-5a + 6b) = 6a - 7b,\\ 6 &= 18 - 12 = (-5a + 6b) - (6a - 7b) = -11a + 13b\end{aligned}$$

□

Notice that the numbers in the left column are precisely the remainders computed by the Euclidean Algorithm. With a little care, we can turn this into a nice theorem, the Extended Euclidean Algorithm.

THEOREM 3.3.4 Suppose a and b are integers, not both zero. Then there are integers x and y such that $(a, b) = ax + by$.

Proof. The Euclidean Algorithm proceeds by finding a sequence of remainders, r_1, r_2, r_3, and so on, until one of them is the gcd. We prove by induction that each r_i is a linear combination of a and b. It is most convenient to assume $a > b$ and let $r_0 = a$ and $r_1 = b$. Then r_0 and r_1 are linear combinations of a and b, which is the base of the induction. The repeated step in the Euclidean Algorithm defines r_{n+2} so that $r_n = qr_{n+1} + r_{n+2}$, or $r_{n+2} = r_n - qr_{n+1}$. If r_n and r_{n+1} are linear combinations of a and b (this is the induction hypothesis) then so is r_{n+2}. ■

Exercises 3.3.

1. For the pairs of integers a, b given below, find the gcd g and integers x and y satisfying $g = ax + by$:
   a) $a = 13, b = 32$
   b) $a = 40, b = 148$
   c) $a = 55, b = 300$
2. If p is a prime, and a is a positive integer, describe (a, p).
3. Suppose g is the gcd of a and b. If i and j are integers and $c = ai + bj$, prove $g|c$.
4. Suppose g is the gcd of a and b. If $g|c$, prove that there are integers i and j such that $c = ai + bj$.
5. If $g = (a, b)$ and $x = ab$, prove $g^2|x$.
6. Suppose that $d|a$ and $d|b$. Prove that $d|(a, b)$.
7. Suppose $g > 0$ and x is a multiple of g^2. Show that there are integers a and b such that $(a, b) = g$ and $ab = x$. (Hint: there is an n such that $x = g^2n$; aim for a trivial case remembering that *you* get to define a and b.)
8. Show that there are, in fact, an infinite number of ways of expressing (a, b) as a combination of a and b. (Hint: use one way to generate an infinite number of other possibilities.)
9. In the proof of theorem 3.3.4, suppose that $r_n = x_n a + y_n b$ and $r_{n+1} = x_{n+1}a + y_{n+1}b$, by the induction hypothesis. Write r_{n+2} as an explicit linear combination of a and b, and identify x_{n+2} and y_{n+2}.
10. The Euclidean algorithm works so well that it is difficult to find pairs of numbers that make it take a long time. Find two numbers whose gcd is 1, for which the Euclidean Algorithm takes 10 steps.
11. Prove that $(F_n, F_{n-1}) = 1$ where F_n is the nth Fibonacci number. (See exercise 7 in section 2.4.)
12. Write a computer program to implement the Extended Euclidean Algorithm. That is, given a and b, the program should compute and display $\gcd(a, b)$, x, and y.

3.4 $\mathbb{U}_n$

At the end of section 3.2 we saw that $\mathbb{Z}_n$ is in some arithmetic ways different than $\mathbb{Z}$. Much of this can be traced to the fact that not all elements of $\mathbb{Z}_n$ have multiplicative inverses (you may have noticed that when we discussed simple arithmetic in $\mathbb{Z}_n$ we left out division). Here we develop a "slimmed down" version of $\mathbb{Z}_n$ that behaves nicely with respect to division.

DEFINITION 3.4.1 The integers a and b are **relatively prime** if $(a, b) = 1$. □

THEOREM 3.4.2 Suppose a and b are integers and b is positive. The following statements are equivalent:

a) a and b are relatively prime,

b) there are integers x and y such that $ax + by = 1$,

c) there is an integer x such that $ax \equiv 1 \pmod{b}$.

Proof. (a) $\Rightarrow$ (b) follows from the Extended Euclidean Algorithm. To prove (b) $\Rightarrow$ (c), note that if $ax + by = 1$, then $ax - 1 = (-y)b$, so $ax \equiv 1 \pmod{b}$. Finally, to prove (c) $\Rightarrow$ (a), if $ax \equiv 1 \pmod{b}$, then there is a z such that $ax = bz + 1$, or $ax - bz = 1$. If $g = (a, b)$, then $g|a$ and $g|b$, so $g|ax - bz = 1$, which means $g = 1$. ■

DEFINITION 3.4.3 If n is a positive integer, let $\mathbb{U}_n \subseteq \mathbb{Z}_n$ consist of those $[u]$ such that for some $[v]$, $[u] \cdot [v] = [1]$, namely, those elements of $\mathbb{Z}_n$ that have multiplicative inverses. □

The invertible elements of $\mathbb{Z}_n$ are sometimes called **units**—hence the $\mathbb{U}$. We say $[v]$ is an **inverse** (or **reciprocal**) of $[u]$. If we translate the last result into the language of $\mathbb{Z}_n$ we have the following:

COROLLARY 3.4.4 If n is a positive integer, then $[u] \in \mathbb{U}_n$ if and only if u and n are relatively prime.

Proof. Immediate. ■

EXAMPLE 3.4.5 $\mathbb{U}_5 = \{[1], [2], [3], [4]\}$. [2] and [3] are inverses of each other, while [1] and [4] are their own inverses. □

EXAMPLE 3.4.6 $\mathbb{U}_{14} = \{[1], [3], [5], [9], [11], [13]\}$. [3] and [5] are inverses, as are [9] and [11]; [1] and [13] are their own inverses. □

In these examples it was easy to find an inverse by inspection. In general, this can be done by the Extended Euclidean Algorithm.

EXAMPLE 3.4.7 $[17] \in \mathbb{U}_{37}$. We apply the Extended Euclidean Algorithm to find: $-13 \cdot 17 + 6 \cdot 37 = 1$, so $[-13] = [24]$ is an inverse for $[17]$. □

We have referred to "an" inverse, but there is only one.

THEOREM 3.4.8 If $[u] \in \mathbb{U}_n$ then the inverse of $[u]$ is unique and is also an element of $\mathbb{U}_n$.

Proof. Suppose $[v_1]$ and $[v_2]$ are both inverses of $[u]$. Then

$$[v_1] = [v_1] \cdot [1] = [v_1] \cdot [u] \cdot [v_2] = [1] \cdot [v_2] = [v_2],$$

which implies uniqueness. Observe that if $[u] \cdot [v] = [1]$, then $[v] \cdot [u] = [1]$ so $[v]$ has an inverse, namely $[u]$, and so it is in $\mathbb{U}_n$. ■

We denote the inverse of $[u]$ by $[u]^{-1}$. Note well that this notation only makes sense if $[u] \in \mathbb{U}_n$.

THEOREM 3.4.9 The product of any two elements of $\mathbb{U}_n$ is an element of $\mathbb{U}_n$.

Proof. Suppose $[u_1]$ and $[u_2]$ are in $\mathbb{U}_n$ with inverses $[v_1]$ and $[v_2]$. Then

$$([u_1] \cdot [u_2]) \cdot ([v_1] \cdot [v_2]) = ([u_1] \cdot [v_1]) \cdot ([u_2] \cdot [v_2]) = [1] \cdot [1] = [1],$$

so $[u_1] \cdot [u_2]$ has an inverse, namely $[v_1] \cdot [v_2]$, and so it is in $\mathbb{U}_n$. ■

EXAMPLE 3.4.10 Here is a multiplication table for $\mathbb{U}_9$:

$\times$	[1]	[2]	[4]	[5]	[7]	[8]
[1]	[1]	[2]	[4]	[5]	[7]	[8]
[2]	[2]	[4]	[8]	[1]	[5]	[7]
[4]	[4]	[8]	[7]	[2]	[1]	[5]
[5]	[5]	[1]	[2]	[7]	[8]	[4]
[7]	[7]	[5]	[1]	[8]	[4]	[2]
[8]	[8]	[7]	[5]	[4]	[2]	[1]

□

Notice that every row contains a [1], as it must, allowing us to read off inverses: $[1]^{-1} = [1]$, $[2]^{-1} = [5]$, $[4]^{-1} = [7]$, $[8]^{-1} = [8]$.

In $\mathbb{Z}_n$ we can add, subtract and multiply, but we cannot always divide. Since division by $[x]$ is the same as multiplication by $[x]^{-1}$, in $\mathbb{Z}_n$ we can divide by precisely those elements which are in $\mathbb{U}_n$. Thus, if p is a prime, algebra in $\mathbb{Z}_p$ is much like algebra in $\mathbb{R}$ or $\mathbb{Q}$. (Why?)

Exercises 3.4.

1. Construct multiplication tables for $\mathbb{U}_5$ and $\mathbb{U}_{14}$.

2. Use the Extended Euclidean Algorithm to compute $[u]^{-1}$ in $\mathbb{U}_n$ where

a) $u = 5$, $n = 13$, **b)** $u = 13$, $n = 19$.

3. Using the fact that in $\mathbb{U}_{39}$, $[4]^{-1} = [10]$, find $[16]^{-1}$.

4. Suppose $g = \gcd(a, b)$. Since g divides a, b there are integers a' and b' with $a = a'g$, $b = b'g$. Prove that a' and b' are relatively prime.

5. How many elements are there in $\mathbb{U}_{243}$? $(243 = 3^5)$

6. Suppose n is positive and $n|ab$. If n and a are relatively prime, prove $n|b$. (Hint: in $\mathbb{Z}_n$, $[a] \cdot [b] = [0]$.)

7. If $[u] \in \mathbb{U}_n$, prove that for every $[y] \in \mathbb{U}_n$ there is a unique $[x] \in \mathbb{U}_n$ such that $[u] \cdot [x] = [y]$. Prove the following consequence of this exercise that we will use in a later section: If $[u] \in \mathbb{U}_n$ and if $[a_1], \ldots, [a_k]$ is a list of all the elements of $\mathbb{U}_n$, then so is $[u][a_1], \ldots, [u][a_k]$.

8. Suppose $[u] \in \mathbb{U}_n$.

a) Show that there are distinct positive integers i and j such that $[u]^i = [u]^j$. (Hint: How many elements are in the set $\{[u]^i : i \in \mathbb{N}\}$?)

b) Use part (a) to show that there is a positive integer k such that $[u]^k = [1]$.

c) What is $[u]^{k-1}$?

9. Suppose $[u] \in \mathbb{U}_n$. It is easy to see that $[u]^i[u]^j = [u]^{i+j}$ and $([u]^i)^j = [u]^{ij}$ if i and j are positive integers. Define $[u]^0 = [1]$ and $[u]^{-k} = ([u]^{-1})^k$ if k is a positive integer. Prove that $[u]^{-k} = ([u]^k)^{-1}$ when k is a positive integer, and that $[u]^i[u]^j = [u]^{i+j}$ and $([u]^i)^j = [u]^{ij}$ for all integers i and j.

3.5 The Fundamental Theorem of Arithmetic

We are ready to prove the Fundamental Theorem of Arithmetic. Recall that this is an ancient theorem—it appeared over 2000 years ago in Euclid's *Elements*.

THEOREM 3.5.1 If $n > 1$ is an integer then it can be factored as a product of primes in exactly one way. In other words, in any two factorizations of n into primes, every prime p occurs the same number of times in each factorization.

Proof. We already have seen that n can be factored in at least one way, in theorem 2.4.4, so we need only prove uniqueness. The proof is by contradiction. Suppose, for instance, that p occurs $i \geq 0$ times in one prime factorization of n, but $j > i$ times in a different prime factorization, so

$$p^i p_1 p_2 \cdots p_k = n = p^j q_1 q_2 \cdots q_l,$$

where each p_m and q_m is a prime different from p. Canceling p^i gives

$$p_1 p_2 \cdots p_k = p^{j-i} q_1 q_2 \cdots q_l.$$

Since $(p_m, p) = 1$, $[p_m] \in \mathbb{U}_p$, by corollary 3.4.4, and so $[p_1 p_2 \cdots p_k] \in \mathbb{U}_p$, by theorem 3.4.9. On the other hand, $[p_1 p_2 \cdots p_k] = [p^{j-i} q_1 q_2 \cdots q_l] = [0]$, and since $[0] \notin \mathbb{U}_p$ we have a contradiction. ■

Collecting like primes, this theorem says that any integer $n > 1$ can be expressed uniquely in the form

$$p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$$

where $p_1 < p_2 < \cdots < p_k$ are distinct primes and the e_i are positive integers. Often we wish to compare the prime factorizations of different integers. If we have two (positive) integers, say a and b, the prime factorization of a may use a different set of primes than the prime factorization of b; that is, some prime p may occur in the prime factorization of a but not b (or vice versa). If we wish to use the same set of primes in both factorizations,

we simply include $p^0 = 1$ in the prime factorization of b. For example, if a factors as $2^2 \cdot 3^5 \cdot 7^3$ and b factors as $3^2 \cdot 5^4 \cdot 11^3$, then we can write

$$a = 2^2 \cdot 3^5 \cdot 5^0 \cdot 7^3 \cdot 11^0$$
$$b = 2^0 \cdot 3^2 \cdot 5^4 \cdot 7^0 \cdot 11^3$$

Such representations are not unique, of course, though they are unique except for the primes that appear with exponent 0. When using an expression like $p_1^{e_1} p_2^{e_2} \ldots p_k^{e_k}$, be sure that you make clear whether or not the e_i are positive or merely non-negative; if the latter, remember not to invoke more uniqueness than is justified. Here's a simple but useful theorem that uses this approach.

THEOREM 3.5.2 If a and b are positive integers, $a = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ and $b = p_1^{f_1} p_2^{f_2} \cdots p_k^{f_k}$ (where the p_i are distinct and the e_i and f_i are non-negative), then a divides b if and only if $e_i \leq f_i$ for every i from 1 to k.

Proof. Let $x = p_1^{t_1} p_2^{t_2} \cdots p_k^{t_k}$, so

$$ax = (p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k})(p_1^{t_1} p_2^{t_2} \cdots p_k^{t_k}) = p_1^{e_1+t_1} p_2^{e_2+t_2} \cdots p_k^{e_k+t_k}.$$

Thus, there is an x such that $ax = b$ if and only if there are non-negative integers $t_1, \ldots, t_k$ such that $e_i + t_i = f_i$ for every i from 1 to k. Clearly such t_i exist if and only if $e_i \leq f_i$ for every i from 1 to k. ∎

Exercises 3.5.

1. Let $a = 3^2 \cdot 5 \cdot 7^3 \cdot 13$ and $b = 2^2 \cdot 3^2 \cdot 5^2 \cdot 7^3 \cdot 11 \cdot 13^4$. Show that $a|b$ by finding an x such that $b = ax$.

2. Suppose $a = p_1^{e_1} p_1^{e_2} \cdots p_i^{e_i}$, with $p_1 < p_2 < \cdots < p_i$. Describe conditions on the prime factorization of a that are equivalent to the following statements. Does it make a difference whether the e_i are allowed to be 0?

a) a is even

b) a is odd

c) a is a perfect square

d) a is a perfect cube

e) a is square-free (i.e., the only divisor of a which is a perfect square is 1)

3. Suppose $n > 0$ and n is not a perfect square. Prove that $\sqrt{n}$ is not rational.

4. Prove that if $a > 0$, $b > 0$ and $a^2|b^2$, then $a|b$.

5. If $a = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, how many positive divisors does a have?

6. When does a positive integer a have an odd number of positive divisors?

7. Find the smallest positive integer x such that $2x$ is a perfect square and $3x$ is a perfect cube; prove that it is the smallest.

8. **a)** Show that $10|a^2$ implies $10|a$.

b) What integers n have the property that for all a, $n|a^2$ implies $n|a$?

9. How many zeros are there on the end of (1000!)?

10. In the proof of the Fundamental Theorem, we made the following statement: "Since x is a product of primes other than p, $[x] \in \mathbb{U}_p$, by corollary 3.4.4 and theorem 3.4.9." But theorem 3.4.9 concerns the product of only two elements of $\mathbb{U}_n$. Prove, by induction, that $\mathbb{U}_n$ is closed under an arbitrary number of multiplications.

3.6 The GCD and the LCM

Many ideas in number theory can be interpreted profitably in terms of prime factorizations. For example, the gcd of two numbers depends directly and simply on their factorizations, and this approach gives us significant new information.

THEOREM 3.6.1 Suppose n and m are positive integers, with prime factorizations $n = p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$ and $m = p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k}$, where the p_i are distinct and the exponents are non-negative. Then:

a) A positive integer $d = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ is a common divisor of n and m if and only if

$$e_1 \leq \min\{i_1, j_1\},\ e_2 \leq \min\{i_2, j_2\}, \ldots,\ e_k \leq \min\{i_k, j_k\}.$$

b) $(n, m) = p_1^{\min\{i_1, j_1\}} p_2^{\min\{i_2, j_2\}} \cdots p_k^{\min\{i_k, j_k\}}$.

c) Any common factor of n and m divides (n, m).

Proof. Although the notation is admittedly rather formidable, this result is a simple consequence of theorem 3.5.2, which says that one number divides another if and only if the primes in the factorization of the first are present to lower powers than those in the second. So if d divides both n and m then any prime in its factorization must occur less often than it occurs in either the factorization of n or the factorization of m; this is just what (a) says. Now to get the largest possible common factor, we clearly should choose the largest exponent possible for each prime, which is exactly what (b) says. Finally, (c) follows immediately from (a), (b) and theorem 3.5.2. ■

You may have used the algorithm implied by (b) to compute gcd's in the past. Recall that we have seen (c) before, in exercise 6 of section 3.3.

We define now another number which is 'dual' to the gcd. If m and n are positive numbers, we let $[m, n]$ or $\text{lcm}(m, n)$ denote their **least common multiple**, or "lcm", that is, the smallest positive number that is a multiple of both m and n. There is an obvious similarity between theorem 3.6.1 and the following result.

THEOREM 3.6.2 Suppose n and m are positive integers, with prime factorizations $n = p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$ and $m = p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k}$, where the p_i are distinct and the exponents are non-negative. Then:

a) A positive integer $s = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ is a common multiple of n and m if and only if

$$e_1 \geq \max\{i_1, j_1\},\ e_2 \geq \max\{i_2, j_2\}, \ldots,\ e_k \geq \max\{i_k, j_k\}.$$

b) $[n, m] = p_1^{\max\{i_1, j_1\}} p_2^{\max\{i_2, j_2\}} \cdots p_k^{\max\{i_k, j_k\}}$.

c) $[n, m]$ divides any common multiple of n and m.

Proof. Entirely analogous to the proof of theorem 3.6.1. ■

The following consequence of the last two results is perhaps a bit surprising, although it is not hard to prove.

THEOREM 3.6.3 If n and m are positive integers, then

$$(n, m) \cdot [n, m] = nm.$$

Proof. Suppose $n = p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$ and $m = p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k}$; then

$$nm = p_1^{i_1 + j_k} p_2^{i_2 + j_2} \cdots p_k^{i_k + j_k}$$

and

$$(n, m) \cdot [n, m] = p_1^{\min\{i_1, j_1\} + \max\{i_1, j_1\}} p_2^{\min\{i_2, j_2\} + \max\{i_2, j_2\}} \cdots \cdots p_k^{\min\{i_k, j_k\} + \max\{i_k, j_k\}}.$$

These are equal because $i + j = \min\{i, j\} + \max\{i, j\}$. ■

EXAMPLE 3.6.4 Let $m = 4$ and $n = 6$. Then it is easy to see that $(4, 6) = 2$ and $[4, 6] = 12$, and $2 \cdot 12 = 4 \cdot 6$. □

Exercises 3.6.

1. If $a = 2 \cdot 3^2 \cdot 7^3 \cdot 13^4$ and $b = 2^5 \cdot 3^2 \cdot 5 \cdot 7^2 \cdot 11^3$, find (a, b), $[a, b]$ and ab.

2. Suppose p is a prime, $(a, p^3) = p^2$ and $(b, p^3) = p$. Find $(a + b, p^3)$.

3. Suppose $a > 0$, $(a, 42) = 6$ and $[a, 42] = 420$. Find a.

4. Show $(na, nb) = n(a, b)$ and $[na, nb] = n[a, b]$.

5. Show that a and b are relatively prime if and only if $ab = [a, b]$.

6. Show $(a^2, b^2) = (a, b)^2$.

7. Suppose g and L are positive integers. Show that there are integers a and b with $(a, b) = g$ and $[a, b] = L$ if and only if $g | L$.

8. Suppose $[a, b] = a^2$. What can you conclude?

9. Prove theorem 3.6.2.

10. In the proof of 3.6.3 we said, "These are equal because $i+j = \min\{i, j\}+\max\{i, j\}$." Explain carefully why this is true; pay attention to the case that $i = j$.

3.7 The Chinese Remainder Theorem

We have taken some pains to note that $\mathbb{Z}_n$ is not a subset of $\mathbb{Z}$, and in particular that $\mathbb{Z}_n = \{[0], [1], \ldots, [n-1]\}$ is not the same as $\{0, 1, \ldots, n-1\}$. The two sets certainly are closely related, however; $[a] = [b]$ if and only if a and b have the same remainder when divided by n, and the numbers in $\{0, 1, \ldots, n-1\}$ are precisely all possible remainders—that's exactly why we chose them to be the "standard" representatives when we write $\mathbb{Z}_n$. This is all by way of pointing out that anything involving $\mathbb{Z}_n$ can be thought of as being "about" remainders. The principal result in this section, the Chinese Remainder Theorem, is an interesting fact about the relationship between $\mathbb{Z}_n$ for different values of n.

NB (That's Latin for "Pay attention!"): In the next two sections (,) denotes an ordered pair, not a gcd.

EXAMPLE 3.7.1 Both $\mathbb{Z}_{12}$ and $\mathbb{Z}_3 \times \mathbb{Z}_4$ have 12 elements. In fact, there is a natural way to associate the elements of $\mathbb{Z}_{12}$ and $\mathbb{Z}_3 \times \mathbb{Z}_4$ given by the following:

$$
\begin{array}{lcllcl}
[0] & \leftrightarrow & ([0],[0]) & [6] & \leftrightarrow & ([6],[6]) = ([0],[2]) \\
[1] & \leftrightarrow & ([1],[1]) & [7] & \leftrightarrow & ([7],[7]) = ([1],[3]) \\
[2] & \leftrightarrow & ([2],[2]) & [8] & \leftrightarrow & ([8],[8]) = ([2],[0]) \\
[3] & \leftrightarrow & ([3],[3]) = ([0],[3]) & [9] & \leftrightarrow & ([9],[9]) = ([0],[1]) \\
[4] & \leftrightarrow & ([4],[4]) = ([1],[0]) & [10] & \leftrightarrow & ([10],[10]) = ([1],[2]) \\
[5] & \leftrightarrow & ([5],[5]) = ([2],[1]) & [11] & \leftrightarrow & ([11],[11]) = ([2],[3])
\end{array}
$$

The relationship used here, $[x] \leftrightarrow ([x], [x])$, is about the simplest one could imagine, and this is one of those happy circumstances in which the simple, obvious choice is the one that works. Be sure you understand that the whole point of this example is to notice that every pair in $\mathbb{Z}_3 \times \mathbb{Z}_4$ appears *exactly* once. When two sets are paired up in this way, so that every element of each set appears in exactly one pair, we say that there is a **one-to-one correspondence** between the sets. Note also that in the expression $[x] \leftrightarrow ([x], [x])$, the symbol $[x]$ means three different things in the three places it appears, namely, $[x] \in \mathbb{Z}_{12}$, $[x] \in \mathbb{Z}_3$, and $[x] \in \mathbb{Z}_4$, respectively. This is an example of a general phenomenon. □

THEOREM 3.7.2 Chinese Remainder Theorem Suppose $n = ab$, with a and b relatively prime. For $x = 0, 1, \ldots, n-1$, associate $[x] \in \mathbb{Z}_n$ with $([x], [x]) \in \mathbb{Z}_a \times \mathbb{Z}_b$ (note that the symbol $[x]$ means different things in $\mathbb{Z}_n$, $\mathbb{Z}_a$ and $\mathbb{Z}_b$). This gives a one-to-one correspondence between $\mathbb{Z}_n$ and $\mathbb{Z}_a \times \mathbb{Z}_b$.

Proof. Observe that the two sets have the same number of elements. If we can show that associating $[x]$ with $([x],[x])$ does not associate any two distinct elements of $\mathbb{Z}_n$ with the same ordered pair in $\mathbb{Z}_a \times \mathbb{Z}_b$, then every element of $\mathbb{Z}_n$ will have to be associated with exactly one element of $\mathbb{Z}_a \times \mathbb{Z}_b$, and vice versa.

Suppose that $[x_1]$ and $[x_2]$ are assigned to the same pair in $\mathbb{Z}_a \times \mathbb{Z}_b$. We wish to show that $[x_1] = [x_2]$. We have

$$x_1 \equiv x_2 \pmod{a} \quad \text{and} \quad x_1 \equiv x_2 \pmod{b},$$

in other words, both a and b must divide $x_1 - x_2$. Since a and b are relatively prime, their product, n, also divides $x_1 - x_2$. (See exercise 3.) That is, $x_1 \equiv x_2 \pmod{n}$ so $[x_1] = [x_2]$. ■

EXAMPLE 3.7.3 The theorem produces a correspondence between $\mathbb{Z}_{168}$ and $\mathbb{Z}_8 \times \mathbb{Z}_{21}$. This correspondence, for example, takes $[97]$ to $([97],[97]) = ([1],[13])$. □

Given an element of $\mathbb{Z}_n$, it is easy to find the corresponding element of $\mathbb{Z}_a \times \mathbb{Z}_b$: simply reduce modulo a and b. Is there a way to reverse this? In other words, given $([y],[z]) \in \mathbb{Z}_a \times \mathbb{Z}_b$ can we find the element of $\mathbb{Z}_n$ to which it corresponds? In fact, using the ubiquitous Extended Euclidean Algorithm it is easy.

EXAMPLE 3.7.4 Which element of $\mathbb{Z}_{168}$ corresponds to the pair $([7],[5]) \in \mathbb{Z}_8 \times \mathbb{Z}_{21}$? If we apply the Extended Euclidean Algorithm to 8 and 21, we get $1 = 8 \cdot 8 + (-3) \cdot 21$. Note that $8 \cdot 8 = 64$ is congruent to 0 mod 8 and 1 mod 21, and $(-3) \cdot 21 = -63$ is congruent to 1 mod 8 and 0 mod 21. Therefore

$$5 \cdot 64 + 7 \cdot (-63) \equiv 5 \cdot 0 + 7 \cdot 1 = 7 \pmod{8},$$
$$5 \cdot 64 + 7 \cdot (-63) \equiv 5 \cdot 1 + 7 \cdot 0 = 5 \pmod{21},$$

so

$$5 \cdot 64 + 7 \cdot (-63) = -121 \equiv 47 \pmod{168}$$

works, that is, $[47] \leftrightarrow ([7],[5])$. □

This last example can be phrased in a somewhat different way. Given 7 and 5, we are asking whether the two simultaneous congruences $x \equiv 7 \pmod{8}$ and $x \equiv 5 \pmod{21}$ can be solved, that is, is there an integer x that has remainder 7 when divided by 8 and remainder 5 when divided by 21. This makes the name "Chinese Remainder Theorem" seem a little more appropriate.

The Chinese Remainder Theorem is a useful tool in number theory (we'll use it in section 3.8), and also has proved useful in the study and development of modern cryptographic systems.

Exercises 3.7.

1. Construct the correspondences between the indicated sets.

a) $\mathbb{Z}_{10}$ and $\mathbb{Z}_2 \times \mathbb{Z}_5$ **b)** $\mathbb{Z}_{15}$ and $\mathbb{Z}_3 \times \mathbb{Z}_5$

2. Given the following values of a, b, and the element $([y],[z])$ of $\mathbb{Z}_a \times \mathbb{Z}_b$, use the Extended Euclidean Algorithm to find the corresponding element of $\mathbb{Z}_{ab}$. In other words, solve the simultaneous congruences $x \equiv y \pmod{a}$ and $x \equiv z \pmod{b}$, as in Example 3.7.4.

a) $a = 11$, $b = 19$, $([5],[12])$ **b)** $a = 9$, $b = 16$, $([1],[4])$

3. The following fact was used in the proof of the Chinese Remainder Theorem: If a and b both divide c, and a and b are relatively prime, then ab divides c. Prove the statement.

4. Suppose in the correspondence between $\mathbb{Z}_{150}$ and $\mathbb{Z}_{25} \times \mathbb{Z}_6$ that $[x]$ corresponds to $([10],[5])$ and $[y]$ corresponds to $([21],[4])$. What does $[x+y]$ correspond to? What does $[xy]$ correspond to?

5. Suppose $n = ab$ where a and b are *not* relatively prime. We can still associate $[x] \in \mathbb{Z}_n$ with $([x],[x]) \in \mathbb{Z}_a \times \mathbb{Z}_b$. Show that this fails to be a one-to-one correspondence. (Hint: Let L be the least common multiple of a and b. Compare $[0]$ and $[L]$.)

6. Observe that there are one-to-one correspondences between $\mathbb{Z}_{60}$ and $\mathbb{Z}_4 \times \mathbb{Z}_{15}$ and between $\mathbb{Z}_4 \times \mathbb{Z}_{15}$ and $\mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_5$.

a) What triples in $\mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_5$ correspond to the following elements of $\mathbb{Z}_{60}$?

(i) $[28]$ (ii) $[59]$ (iii) $[47]$

b) State a theorem generalizing the example in part (a) (you need not prove it).

7. Show that $\mathbb{U}_n = \{[x] \in \mathbb{Z}_n : [x] \neq [0]\}$ if and only if n is a prime.

8. Using example 3.7.4 as a guide, give an alternate proof of Theorem 3.7.2, by showing that for every $([x],[y]) \in \mathbb{Z}_a \times \mathbb{Z}_b$ there is a $[z] \in \mathbb{Z}_n$ such that $([z],[z]) = ([x],[y])$.

3.8 The Euler Phi Function

When something is known about $\mathbb{Z}_n$, it is frequently fruitful to ask whether something comparable applies to $\mathbb{U}_n$. Here we look at $\mathbb{U}_n$ in the context of the previous section. To aid the investigation, we introduce a new quantity, the **Euler phi function**, written $\phi(n)$, for positive integers n.

DEFINITION 3.8.1 $\phi(n)$ is the number of non-negative integers less than n that are relatively prime to n. In other words, if $n > 1$ then $\phi(n)$ is the number of elements in $\mathbb{U}_n$, and $\phi(1) = 1$. □

EXAMPLE 3.8.2 You can verify readily that $\phi(2) = 1$, $\phi(4) = 2$, $\phi(12) = 4$ and $\phi(15) = 8$. □

EXAMPLE 3.8.3 If p is a prime, then $\phi(p) = p - 1$, because $1, 2, \ldots, p-1$ are all relatively prime to p, and 0 is not. □

For any number n, $\phi(n)$ turns out to have a remarkably simple form; that is, there is a simple formula that gives the value of $\phi(n)$. We've already seen how simple it is for primes. As is typical of many results in number theory, we will work our way gradually to any n, looking next at powers of a single prime.

THEOREM 3.8.4 If p is a prime and a is a positive integer, then

$$\phi(p^a) = p^a - p^{a-1}$$

Proof. We want to calculate the number of non-negative integers less than $n = p^a$ that are relatively prime to n. As in many cases, it turns out to be easier to calculate the number that are *not* relatively prime to n, and subtract from the total. List the non-negative integers less than p^a: 0, 1, 2, ..., $p^a - 1$; there are p^a of them. The numbers that have a common factor with p^a (namely, the ones that are not relatively prime to n) are the multiples of p: 0, p, $2p$, ..., that is, every pth number. There are thus $p^a/p = p^{a-1}$ numbers in this list, so $\phi(p^a) = p^a - p^{a-1}$. ■

EXAMPLE 3.8.5 $\phi(32) = 32 - 16 = 16$, $\phi(125) = 125 - 25 = 100$. □

Now we want to extend our formula to handle any positive integer n. Consider an example first:

EXAMPLE 3.8.6 Since

$$\begin{aligned}\mathbb{U}_{20} &= \{[1],[3],[7],[9],[11],[13],[17],[19]\},\\ \mathbb{U}_4 &= \{[1],[3]\},\\ \mathbb{U}_5 &= \{[1],[2],[3],[4]\},\end{aligned}$$

both $\mathbb{U}_{20}$ and $\mathbb{U}_4 \times \mathbb{U}_5$ have 8 elements. In fact, the correspondence discussed in the Chinese Remainder Theorem between $\mathbb{Z}_{20}$ and $\mathbb{Z}_4 \times \mathbb{Z}_5$ is also a 1-1 correspondence between $\mathbb{U}_{20}$ and $\mathbb{U}_4 \times \mathbb{U}_5$:

$$\begin{array}{lcllcl} [1] & \leftrightarrow & ([1],[1]) & [11] & \leftrightarrow & ([3],[1]) \\ [3] & \leftrightarrow & ([3],[3]) & [13] & \leftrightarrow & ([1],[3]) \\ [7] & \leftrightarrow & ([3],[2]) & [17] & \leftrightarrow & ([1],[2]) \\ [9] & \leftrightarrow & ([1],[4]) & [19] & \leftrightarrow & ([3],[4]) \end{array}$$

□

Using the Chinese Remainder Theorem we can prove that this is true in general.

THEOREM 3.8.7 If a and b are relatively prime and $n = ab$, then $\phi(n) = \phi(a)\phi(b)$.

Proof. We want to prove that $|\mathbb{U}_n| = |\mathbb{U}_a| \cdot |\mathbb{U}_b|$. As indicated in the example, we will actually prove more, by exhibiting a one to one correspondence between the elements of

$\mathbb{U}_n$ and $\mathbb{U}_a \times \mathbb{U}_b$. We already have a one to one correspondence between the elements of $\mathbb{Z}_n$ and $\mathbb{Z}_a \times \mathbb{Z}_b$. Again as indicated by the example, we just have to prove that this same correspondence works for $\mathbb{U}_n$ and $\mathbb{U}_a \times \mathbb{U}_b$. That is, we already know how to associate any $[x]$ with a pair $([x],[x])$; we just need to know that $[x] \in \mathbb{U}_n$ if and only if $([x],[x]) \in \mathbb{U}_a \times \mathbb{U}_b$. After a long-winded build up, here's the proof: $[x]$ is in $\mathbb{U}_n$ if and only if $(x,n) = 1$ if and only if $(x,a) = 1$ and $(x,b) = 1$ if and only if $([x],[x]) \in \mathbb{U}_a \times \mathbb{U}_b$. ■

COROLLARY 3.8.8 Suppose $n = ab$, with a and b relatively prime. For $x = 0, 1, \ldots, n-1$, if $[x] \in \mathbb{U}_n$, associate $[x]$ with $([x],[x]) \in \mathbb{Z}_a \times \mathbb{Z}_b$. This gives a one-to-one correspondence between $\mathbb{U}_n$ and $\mathbb{U}_a \times \mathbb{U}_b$.

Proof. We proved this already in the proof of the previous theorem, but it deserves its own statement. ■

Now we know enough to compute $\phi(n)$ for any n.

EXAMPLE 3.8.9 $\phi(200) = \phi(25)\phi(8) = (25-5)(8-4) = 80.$ □

EXAMPLE 3.8.10 $\phi(2^3 3^4 7^2) = \phi(2^3)\phi(3^4 7^2) = \phi(2^3)\phi(3^4)\phi(7^2) = (2^3 - 2^2)(3^4 - 3^3)(7^2 - 7)$

□

We can express this as a formula once and for all:

THEOREM 3.8.11 If n is a positive integer with prime factorization $p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$, then

$$\phi(n) = (p_1^{e_1} - p_1^{e_1 - 1})(p_2^{e_2} - p_2^{e_2 - 1}) \cdots (p_k^{e_k} - p_k^{e_k - 1}).$$

Proof. The proof by induction is left as an exercise. ■

Leonhard Euler. Euler (pronounced "oiler") was born in Basel in 1707 and died in 1783, following a life of stunningly prolific mathematical work. His complete bibliography runs to nearly 900 entries; his research amounted to some 800 pages a year over the whole of his career. He continued doing research right up until his sudden death while relaxing with a cup of tea. For almost all of the last 17 years of his life he was totally blind.

The breadth of Euler's knowledge may be as impressive as the depth of his mathematical work. He had a great facility with languages, and studied theology, medicine,

astronomy and physics. His first appointment was in medicine at the recently established St. Petersburg Academy. On the day that he arrived in Russia, the academy's patron, Catherine I, died, and the academy itself just managed to survive the transfer of power to the new regime. In the process, Euler ended up in the chair of natural philosophy instead of medicine.

Euler is best remembered for his contributions to analysis and number theory, especially for his use of infinite processes of various kinds (infinite sums and products, continued fractions), and for establishing much of the modern notation of mathematics. Euler originated the use of e for the base of the natural logarithms and i for $\sqrt{-1}$; the symbol π has been found in a book published in 1706, but it was Euler's adoption of the symbol, in 1737, that made it standard. He was also responsible for the use of $\sum$ to represent a sum, and for the modern notation for a function, $f(x)$.

Euler's greatest contribution to mathematics was the development of techniques for dealing with infinite operations. In the process, he established what has ever since been called the field of **analysis**, which includes and extends the differential and integral calculus of Newton and Leibniz. For example, by treating the familiar functions $\sin x$, $\cos x$ and e^x analytically (as infinite series), Euler could easily establish identities that became fundamental tools in analysis. One such is the well-known $e^{ix} = \cos x + i \sin x$; substituting $x = \pi$ gives $e^{i\pi} = -1$ or $e^{i\pi} + 1 = 0$, a remarkable equation containing perhaps the five most important constants in analysis.

Euler used infinite series to establish and exploit some remarkable connections between analysis and number theory. Many talented mathematicians before Euler had failed to discover the value of the sum of the reciprocals of the squares: $1^{-2} + 2^{-2} + 3^{-2} + \cdots$. Using the infinite series for $\sin x$, and assuming that it behaved like a finite polynomial, Euler showed that the sum is $\pi^2/6$. Euler's uncritical application of ordinary algebra to infinite series occasionally led him into trouble, but his results were overwhelmingly correct, and were later justified by more careful techniques as the need for increased rigor in mathematical arguments became apparent. We'll see Euler's name more than once in the remainder of the chapter.

The information here is taken from *A History of Mathematics*, by Carl Boyer, New York: John Wiley & Sons, 1968.

Exercises 3.8.

1. Construct the correspondence between

a) $\mathbb{U}_{21}$ and $\mathbb{U}_3 \times \mathbb{U}_7$ **b)** $\mathbb{U}_{30}$ and $\mathbb{U}_5 \times \mathbb{U}_6$

2. Given the following values of a, b and the element $([y], [z])$ of $\mathbb{U}_a \times \mathbb{U}_b$, use the Euclidean Algorithm to find the corresponding element of $\mathbb{U}_n$.

a) $a = 7$, $b = 11$, $([4], [9])$ **b)** $a = 12$, $b = 17$, $([11], [2])$

3. Compute the following:

a) $\phi(512)$ **b)** $\phi(9,000)$

c) $\phi(2^3 \cdot 5^2 \cdot 7^5 \cdot 11^3)$

4. Suppose in the correspondence between $\mathbb{U}_{175}$ and $\mathbb{U}_{25} \times \mathbb{U}_7$ that $[x]$ corresponds to $([13],[2])$. What does $[x]^2$ correspond to? What does $[x]^{-1}$ correspond to?

5. The divisors of 6 are 1, 2, 3, 6. Observe that

$$\phi(1) + \phi(2) + \phi(3) + \phi(6) = 1 + 1 + 2 + 2 = 6.$$

Perform a similar computation with 6 replaced by 10.

6. Find all a such that $\phi(a) = 6$.

7. If $a|b$, prove $\phi(a)|\phi(b)$.

8. What primes can be expressed in the form $\phi(n)$ for some n?

9. Prove that $\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$; the product is over all primes p that divide n.

10. Prove Theorem 3.8.11.

11. Find all n such that $\phi(n)$ is odd, and prove that you have found all such n.

12. In the proof of theorem 3.8.7, we claimed that if $n = ab$ then $(x,n) = 1$ if and only if $(x,a) = 1$ and $(x,b) = 1$. Prove this.

3.9 The Phi Function—Continued

The phi function is a useful tool, but it is also interesting in its own right. Problem 5 in section 3.8 suggested an intriguing identity; it's true in general, and we'll prove it.

Suppose $n > 1$. If d is a positive divisor of n, then there is a positive integer e such that $n = de$. Observe that as d varies over all the positive divisors of n, so does e.

EXAMPLE 3.9.1 If $n = 12$, then as d runs through the list 1, 2, 3, 4, 6, 12, e takes on the values 12, 6, 4, 3, 2, 1, respectively. □

Suppose that $n = de$. Let $G_e = \{x : 0 \le x < n \text{ and } (x,n) = e\}$, that is, G_e consists of all numbers whose gcd with n is e. Every $x \in \{0, 1, \ldots, n-1\}$ is contained in G_e for exactly one divisor e of n, because (x,n) is a divisor of n. In other words, the collection of all the sets G_e as e runs through the divisors of n **partitions** the set $\{0, 1, \ldots, n-1\}$.

EXAMPLE 3.9.2 If $n = 12$, then

$$G_1 = \{1,5,7,11\}, \quad G_2 = \{2,10\}, \quad G_3 = \{3,9\},$$
$$G_4 = \{4,8\}, \quad G_6 = \{6\}, \quad G_{12} = \{0\}.$$

If $n = 15$, then

$$G_1 = \{1,2,4,7,8,11,13,14\}, \quad G_3 = \{3,6,9,12\},$$
$$G_5 = \{5,10\}, \quad G_{15} = \{0\}.$$

□

Notice that each G_e consists of some multiples of e. This is hardly a surprise, since it follows immediately from the definition of G_e. Nevertheless, this simple fact can be exploited in a surprising way. Form a new collection of sets by factoring e out of the elements of G_e. For the example above, this gives the following two lists of sets:

$$\{1,5,7,11\}, \quad \{1,5\}, \quad \{1,3\}, \quad \{1,2\}, \quad \{1\}, \quad \{0\},$$

and

$$\{1,2,4,7,8,11,13,14\}, \quad \{1,2,3,4\}, \quad \{1,2\}, \quad \{0\}.$$

Now, it's not immediately obvious, but with the exception of $\{0\}$, every one of these sets is "almost" a $\mathbb{U}_d$ (what's missing is the [ ] around each element): $\mathbb{U}_{12}$, $\mathbb{U}_6$, $\mathbb{U}_4$, $\mathbb{U}_3$, $\mathbb{U}_2$, $\mathbb{U}_{15}$, $\mathbb{U}_5$, and $\mathbb{U}_3$, respectively. Moreover, notice that the subscript d in every case is n/e. This leads us to define

$$R_d = \{y : 0 \leq y < d \text{ and } (y,d) = 1\}.$$

Notice that this definition makes sense even for $d = 1$, namely, $R_1 = \{0\}$, and in every case, the number of elements in R_d is $\phi(d)$ (because R_d is almost $\mathbb{U}_d$—missing the [ ] again). Based on these examples, it's natural to speculate that multiplication by e sets up a one-to-one correspondence between R_d and G_e (assuming that $n = de$).

LEMMA 3.9.3 If $n = de$, then G_e and R_d have the same number of elements. In fact, a 1-1 correspondence between the two sets is obtained by multiplying every element of R_d by e. As a result, there are $\phi(d)$ elements in G_e.

Proof. Note that

$$0 \leq y < d \quad \text{iff} \quad 0 \leq ey < ed \quad \text{iff} \quad 0 \leq ey < n,$$

and

$$(y,d) = 1 \quad \text{iff} \quad (ey, ed) = e \quad \text{iff} \quad (ey, n) = e.$$

(See exercise 4 of section 3.6.) Together these facts imply that $y \in R_d$ if and only if $ey \in G_e$. ■

Now we can prove the theorem suggested by problem 5 in section 3.8.

THEOREM 3.9.4 Suppose n is a positive integer. Then

$$n = \sum_{0<d|n} \phi(d).$$

Proof. Since each of the n numbers from 0 to $n-1$ is in exactly one G_e,

$$n = \sum_{0<e|n} |G_e| = \sum_{0<e|n} \phi(d) = \sum_{0<d|n} \phi(d).$$

The last equality is valid because as e ranges over the positive divisors of n, so does d, that is, the two sums are exactly the same, but they add up the terms in opposite orders. ■

EXAMPLE 3.9.5 If $n = 20$,

$$\phi(1) + \phi(2) + \phi(4) + \phi(5) + \phi(10) + \phi(20) = 1 + 1 + 2 + 4 + 4 + 8 = 20.$$

If $n = 33$,

$$\phi(1) + \phi(3) + \phi(11) + \phi(33) = 1 + 2 + 10 + 20 = 33.$$

□

Exercises 3.9.

1. For each given n, find the sets R_d, G_e and verify Theorem 3.9.4.
   a) $n = 14$ **b)** $n = 18$
   c) $n = 24$
2. Give a direct proof of Theorem 3.9.4 for the case that $n = p^a$ (p prime); use Theorem 3.8.4.
3. Give a direct proof of Theorem 3.9.4 for the case $n = pq$, where p and q are distinct primes.
4. Show that if n is positive and $(a, n) = g$, then $\exists x(ax \equiv b \pmod{n})$ iff $g|b$.

Suppose we try to find all solutions to

$$ax \equiv b \pmod{n}.$$

Let $g = (a, n)$. If $g \nmid b$, then there are no solutions, by the last problem. Otherwise, $a = rg$, $b = sg$, and $n = n'g$. So by exercise 8 of section 3.1,

$$rx \equiv s \pmod{n'}.$$

By exercise 4 of section 3.4, r and n' are relatively prime. So by Theorem 3.4.2 there is a t such that $tr \equiv 1 \pmod{n'}$, and

$$x \equiv trx \equiv ts \pmod{n'}.$$

So if x is a solution to $ax \equiv b \pmod{n}$, then $x \equiv ts \pmod{n'}$. In fact, every such x really is a solution: Suppose $x = n'q + ts$ for some q. Then

$$\begin{aligned} ax = an'q + ats &= rgn'q + rgts \\ &= rnq + gs(n'k + 1) \\ &= rnq + gn'ks + gs \\ &= rnq + nks + b, \end{aligned}$$

so $ax \equiv b \pmod{n}$.

Example: $12x \equiv 10 \pmod{28}$ has no solutions since $(12, 28) = 4 \nmid 10$. However, if we consider a slightly different problem,

$$12x \equiv 8 \pmod{28},$$

we can reduce it to

$$3x \equiv 2 \pmod{7}.$$

Note that $3 \cdot 5 \equiv 1 \pmod{7}$, so the answer is

$$x \equiv 10 \equiv 3 \pmod{7},$$

i.e., $x \in \{\ldots, -11, -4, 3, 10, \ldots\}$.

5. Solve the following congruences:

a) $30x \equiv 24 \pmod{72}$

b) $30x \equiv 32 \pmod{72}$

c) $66x \equiv 15 \pmod{159}$

6. Suppose $n = p_1^{e_1} \cdots p_i^{e_i}$, where $e_1, \ldots, e_i$ are positive; $m = p_1 \cdots p_i$; and $[x] \in \mathbb{Z}_n$. Show that there is a positive integer k such that $[x]^k = [0]$ iff $m|x$.

The subset of $\mathbb{Z}_n$ consisting of elements $[x]$ such that $[x]^k = [0]$ for some k, is called the **radical** of $\mathbb{Z}_n$.

7. If $[x]$ is in the radical of $\mathbb{Z}_n$, show that $[1 - x] \in \mathbb{U}_n$.

3.10 Wilson's Theorem and Euler's Theorem

The defining characteristic of $\mathbb{U}_n$ is that every element has a unique multiplicative inverse. It is quite possible for an element of $\mathbb{U}_n$ to be its own inverse; for example, in $\mathbb{U}_{12}$, $[1]^2 = [11]^2 = [5]^2 = [7]^2 = [1]$. This stands in contrast to arithmetic in $\mathbb{Z}$ or $\mathbb{R}$, where the only solutions to $x^2 = 1$ are ± 1. If n is prime, then this familiar fact is true in $\mathbb{U}_n$ as well.

THEOREM 3.10.1 If p is a prime, the only elements of $\mathbb{U}_p$ which are their own inverses are $[1]$ and $[p - 1] = [-1]$.

Proof. Note that $[n]$ is its own inverse if and only if $[n^2] = [n]^2 = [1]$ if and only if $n^2 \equiv 1 \pmod{p}$ if and only if $p|(n^2 - 1) = (n - 1)(n + 1)$. This is true if and only if $p|(n - 1)$ or $p|(n + 1)$. In the first case, $n \equiv 1 \pmod{p}$, i.e., $[n] = [1]$. In the second case, $n \equiv -1 \equiv p - 1 \pmod{p}$, i.e., $[n] = [p - 1]$. ∎

If p is prime, $\mathbb{U}_p = \{[1], [2], \ldots, [p - 1]\}$. The elements $[2]$, $[3]$, $\ldots$, $[p - 2]$ all have inverses different from themselves, so it must be possible to pair up each element in this list with its inverse from the list. This means that if we multiply all of $[2]$, $[3]$, $\ldots$, $[p - 2]$ together, we must get $[1]$.

EXAMPLE 3.10.2 If $p = 11$,

$$\begin{aligned}[2] \cdot [3] \cdot [4] \cdot [5] \cdot [6] \cdot [7] \cdot [8] \cdot [9] &= ([2] \cdot [6])([3] \cdot [4])([5] \cdot [9])([7] \cdot [8]) \\ &= [1] \cdot [1] \cdot [1] \cdot [1] = [1].\end{aligned}$$

□

This observation suggests the following, called **Wilson's Theorem**:

THEOREM 3.10.3 If $p > 1$ then p is prime iff $(p-1)! \equiv -1 \pmod{p}$.

Proof. If p is prime, $[(p-1)!] = [p-1] \cdot ([p-2] \cdots [2]) \cdot [1] = [p-1] \cdot [1] \cdot [1] = [p-1]$, and this means that $(p-1)! \equiv p-1 \equiv -1 \pmod{p}$. The other direction is left as an exercise. ■

EXAMPLE 3.10.4 $2! = 2 \equiv -1 \pmod{3}$, $4! = 24 \equiv -1 \pmod{5}$. □

Similar in spirit, and very useful, is **Euler's Theorem**:

THEOREM 3.10.5 If $n > 0$, and u is relatively prime to n, then

$$u^{\phi(n)} \equiv 1 \pmod{n}.$$

Proof. Suppose $k = \phi(n)$ and $[a_1], \ldots, [a_k]$ is a list of the elements of $\mathbb{U}_n$. By exercise 7 of section 3.4, $[u][a_1], \ldots, [u][a_k]$ also is a list of the elements of $\mathbb{U}_n$. Multiplying these together gives

$$[a_1] \cdots [a_k] = [u] \cdot [a_1] \cdots [u] \cdot [a_k] = [u]^k \cdot [a_1] \cdots [a_k].$$

Canceling the $[a_i]$ terms gives $[u]^k = [1]$, i.e., $u^{\phi(n)} \equiv 1 \pmod{n}$. ■

EXAMPLE 3.10.6 If $n = 8$, $u = 3$, then $3^4 = 81 \equiv 1 \pmod{8}$. If $n = 14$, $u = 5$, then $5^6 = 15625 \equiv 1 \pmod{14}$. □

Fermat's Little Theorem follows almost immediately as a special case of Euler's Theorem.

COROLLARY 3.10.7 If p is a prime, then for any a, $a^p \equiv a \pmod{p}$.

Proof. If $p|a$, the result is clear, and if $p \nmid a$ then $(a, p) = 1$, so by theorem 3.10.5, $a^p = a^{p-1}a \equiv 1 \cdot a = a$. ■

EXAMPLE 3.10.8 If $p = 7$ and $a = 3$, then $3^7 = 2187 \equiv 3 \pmod{7}$. □

Exercises 3.10.

1. For $p = 13$ and $p = 19$ find the pairing of elements in $\mathbb{U}_p$ used in example 3.10.2 and (implicitly) in the proof of Wilson's Theorem.
2. The following fact was used in the proof of theorem 3.10.1: If the prime p divides ab, then it divides either a or b. Prove it.
3. Prove the following converse of exercise 2: If $n > 1$ is a number with the property that whenever n divides ab then n divides either a or b, then n is a prime.
4. Use exercises 2 and 3 to prove: $\mathbb{Z}_n$ has the property that $[x] \cdot [y] = [0]$ implies either $[x] = [0]$ or $[y] = [0]$ if and only if n is prime.
5. Prove that if $e > 2$, then $\mathbb{U}_{2^e}$ has an element, other than $[2^e - 1]$ and $[1]$, which is its own inverse. (Hint: in $\mathbb{U}_{32}$, $[17]$ is its own inverse.)
6. Prove that if p is a prime and $e > 0$, and $\mathbb{U}_{p^e}$ has an element, other than $[p^e - 1]$ and $[1]$, which is its own inverse, then $p = 2$. (Hint: If $[x]^2 = [1]$, show $p|(x + 1)$ and $p|(x - 1)$.)
7. Find all n which are the products of their proper positive divisors (e.g., $10 = 2 \cdot 5$ is such a number).
8. Prove that if x is any composite number other than 4, then $(x - 1)! \equiv 0 \pmod{x}$.
9. Verify Euler's Theorem in the following cases:
   a) $u = 3$, $n = 10$ **b)** $u = 5$, $n = 6$ **c)** $u = 2$, $n = 15$
10. Suppose $n > 0$ and u is relatively prime to n.
    a) If $\phi(n)|m$, prove that $u^m \equiv 1 \pmod{n}$.
    b) If m is relatively prime to $\phi(n)$ and $u^m \equiv 1 \pmod{n}$, prove that $u \equiv 1 \pmod{n}$.
11. Finish the proof of Wilson's Theorem, 3.10.3.

3.11 Public Key Cryptography

Until about 1970, cryptography was *private key cryptography:* a secret of some kind (typically a string of letters and numbers) was used both to encrypt and decrypt a message, and so both the sender and receiver had to know the secret key.

For example, all textual messages can be encoded as a sequence of 0s and 1s (bits), and so can the key. Here is a simple way to encrypt such a message: line up the message and the key, and add the bits modulo 2:

message:	110010110001101110101011
key:	101101110010110010001101 0
sum:	011111000011011100100100 1

The sender transmits the sum; the receiver then adds the sum to the key in the same way, and recovers the message. If the message is longer than the key, the key can be repeated as many times as required, though there are techniques that can be used to break this system. The only provably secure method is to use a key as long as the message, and never to reuse a key.

In *public key cryptography*, there are two keys. Suppose Alice wishes to receive encrypted messages; she publishes one of the keys, the public key, and anyone, say Bob, can use it to encrypt a message and send it to her. When Alice gets the encrypted message, she uses the private key to decrypt it and read the original message. If Alice needs to reply to Bob, Bob will publish his own public key, and Alice can use it to encrypt her reply.

We will describe one method of public key cryptography, or *cryptosystem*, called RSA, after Ron Rivest, Adi Shamir and Leonard Adleman.

Alice chooses two prime numbers, p and q, and publishes the product $n = pq$ together with an integer c that is relatively prime to both $p-1$ and $q-1$, that is, to $[p-1, q-1] = L$. Alice also computes $[c]^{-1} = [d]$ in $\mathbb{U}_L$.

To send Alice a message, Bob represents his message as a sequence of integers, each smaller than both p and q. For each of these numbers x, Bob computes x^c, and then the remainder of x^c mod n, so $x^c = nQ + r$. Bob sends r to Alice.

For each number r that Alice receives, she computes r^d mod n; this is the original x. Here's why: Since $[c][d] = [1]$ in $\mathbb{U}_L$, we know that $cd \equiv 1 \pmod{L}$, or $cd = 1 + tL$. Now $r^d \equiv (x^c)^d \pmod{n}$, since $r \equiv x^c \pmod{n}$, and $x^{cd} = x^{1+tL} = x(x^L)^t$. Since $x < p$ and $x < q$, x is relatively prime to n, so $[x] \in \mathbb{U}_n$.

Now consider $[x]^L \in \mathbb{U}_n$. From sections 3.7 and 3.8, we know this is matched with $([x]^L, [x]^L) \in \mathbb{U}_p \times \mathbb{U}_q$. Now

$$L = \frac{(p-1)(q-1)}{(p-1, q-1)} = (p-1)A = (q-1)B,$$

where A and B are integers. Then

$$[x]^L = ([x]^{p-1})^A = ([x]^{\phi(p)})^A = [1]^A = [1]$$

in $\mathbb{U}_p$, using Euler's Theorem (3.10.5), and

$$[x]^L = ([x]^{q-1})^B = ([x]^{\phi(q)})^B = [1]^B = [1]$$

in $\mathbb{U}_q$. Hence $[x]^L$ is paired with $([1], [1])$ in $\mathbb{U}_p \times \mathbb{U}_q$, and since $[1]$ is also paired with $([1], [1])$, $[x]^L = [1]$ in $\mathbb{U}_n$, that is, $x^L \equiv 1 \pmod{n}$.

Now, modulo n, $r^d \equiv x(x^L)^t \equiv x$. Since $x < n$, x is the remainder when r^d is divided by n, that is, $x = r^d \pmod{n}$, as claimed.

Suppose we use $p = 37$, $q = 73$, and $c = 5$. Then $n = 2701$, $d = 29$ and $L = [36, 72] = 72$. Suppose one number in a message is 33. Then Bob computes $33^5 \pmod{2701} = 604$ and sends 604 to Alice. Alice computes $604^{29} \pmod{2701} = 33$, the original number in Bob's message. You can use a Sage worksheet to perform the calculations.

It is possible to break this code by factoring the published number n. While this is in principle easy, there is no known way to factor very large numbers in a reasonable length of

time. In practice each of p and q would be prime numbers with hundreds of digits so that factoring $n = pq$ is not feasible. Also, individual characters are not suitable for encrypting, since then it is possible to attack the code based on the frequency of different characters. Many characters should be grouped together, giving a large number to be encrypted as a block.

Finally, while the necessary operations for encrypting and decrypting can be performed fairly quickly with modern computers, there are good private key cryptosystems that are much faster. Instead of encrypting the entire message with RSA, it can be used to encrypt and exchange a secret key, and this key is then used with another cryptosystem to encrypt the message. This secret key can be generated at random and used only once.

Exercises 3.11.

Use this code to turn symbols into integers: A through Z are represented by 1 through 26, a period is 27, a comma is 28, an exclamation point is 29, a space is 30. (This is used in the Sage worksheet, which you may want to use to do the exercises.)

1. Use $p = 101$, $q = 103$, and $c = 121$ to encrypt "MEET ME AT NOON."
2. Use the values for p, q, and c from the previous problem to decrypt this message: [1403, 6884, 8311, 8311, 7466, 438, 1106, 2589, 7466, 4239, 8311, 1457, 5381].

3.12 Quadratic Reciprocity

The prime numbers, their properties, and their relation to the composite numbers have fascinated mathematicians for thousands of years. Yet it was not until the 1700s that the first really deep result about prime numbers was discovered, by Leonhard Euler. The *Quadratic Reciprocity Theorem* was proved first by Gauss, in the early 1800s, and reproved many times thereafter (at least eight times by Gauss). We conclude our brief study of number theory with a beautiful proof due to the brilliant young mathematician Gotthold Eisenstein, who died tragically young, at 29, of tuberculosis. The proof is similar to one by Gauss, but it replaces a complicated lemma by an ingenious geometrical argument. This is a good place to leave number theory, as it hints at the wonderful but difficult and subtle areas of the subject; we hope it makes you want to explore number theory further. See the bibliography for some starting points.

Suppose p is an odd prime and p does not divide b. Then b is a **quadratic residue** (mod p) if $b \equiv c^2 \pmod{p}$ for some c, and otherwise b is a **quadratic nonresidue**. In other words, a quadratic residue is a "perfect square" in the world of modular arithmetic.

It is easy to see that $x^2 \equiv (p-x)^2 \equiv (-x)^2$ for any x, so at most half of the elements of $\{1, 2, 3, \ldots, p-1\}$ are quadratic residues modulo p. It is also not hard to see that $x^2 \equiv y^2$ implies that $y \equiv \pm x$, so in fact exactly half of the elements of $\{1, 2, 3, \ldots, p-1\}$ are quadratic residues.

It is convenient to have an easy and arithmetic way to identify whether an element is a quadratic residue. It turns out to be most useful to define a function $\mathrm{QR}(b,p) = 1$ if b is a quadratic residue mod p, and $\mathrm{QR}(b,p) = -1$ otherwise. The standard notation for this function is the **Legendre symbol**:

$$\mathrm{QR}(b,p) = \left(\frac{b}{p}\right).$$

We should emphasize here that the notation "QR" is completely non-standard, introduced in the hope that first writing this as a function will make it easier to understand. The Legendre symbol is not an ideal choice, since it looks exactly like a fraction, but as it is the standard notation we will use it throughout the section.

Here is a theorem similar to Wilson's Theorem (3.10.3); in fact, Wilson's Theorem is a simple corollary of this theorem.

THEOREM 3.12.1 If p does not divide b then

$$(p-1)! \equiv -\left(\frac{b}{p}\right) b^{(p-1)/2} \pmod{p}.$$

Proof. Recall that for every $x \in \{1,2,3,\ldots,p-1\}$ there is a unique $y \in \{1,2,3,\ldots,p-1\}$ such that $xy \equiv b$. If b is a quadratic residue then y may be equal to x, but if b is a quadratic nonresidue then $y \neq x$.

Suppose that b is a quadratic nonresidue. Then the numbers 1, 2, ..., $p-1$ can be grouped into $(p-1)/2$ pairs $\{x_i, y_i\}$ with $x_i y_i \equiv b$. Thus

$$(p-1)! = \prod_{i=1}^{(p-1)/2} x_i y_i \equiv b^{(p-1)/2} \pmod{p}.$$

Now suppose that b is a quadratic residue. There are exactly two numbers in $\{1,2,\ldots,p-1\}$, say c and $p-c$, such that $c^2 \equiv (p-c)^2 \equiv b$. The remaining $p-3$ numbers can be paired up as before. Then

$$(p-1)! = c(p-c) \prod_{i=1}^{(p-3)/2} x_i y_i \equiv (pc - c^2) b^{(p-3)/2} \equiv (-b) b^{(p-3)/2} \equiv -b^{(p-1)/2},$$

where all congruences are (mod p). This completes the proof. ∎

COROLLARY 3.12.2 Euler's Criterion If b is not divisible by p then

$$\left(\frac{b}{p}\right) \equiv b^{(p-1)/2} \pmod{p}.$$

Proof. Using Wilson's Theorem 3.10.3 and Theorem 3.12.1,

$$-1 \equiv (p-1)! \equiv -\left(\frac{b}{p}\right) b^{(p-1)/2},$$

so

$$1 \equiv \left(\frac{b}{p}\right) b^{(p-1)/2}.$$

This implies the desired congruence. ■

Now we are ready for the principal result. Until now, we have used the notation "mod" only in a context like $a \equiv b \pmod{n}$. There is another common use of this notation; fortunately it is closely related to the first. If we refer to "a mod n," we mean the remainder when a is divided by n, that is, the unique remainder in $\{0, 1, \ldots, n-1\}$. For example, (25 mod 11) is 3.

THEOREM 3.12.3 Quadratic Reciprocity Theorem If p and q are distinct odd primes, then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{((p-1)/2)((q-1)/2)}.$$

Proof. Let $E = \{2, 4, 6, \ldots, p-1\}$ and $r_e = eq \bmod p$. We claim that

$$\{(-1)^{r_e} r_e \bmod p : e \in E\} = E.$$

First, we note that if r_e is even then $(-1)^{r_e} r_e = r_e$ is even, and if r_e is odd then $(-1)^{r_e} r_e = -r_e$, so $(-1)^{r_e} r_e \bmod p = p - r_e$, which is even. Thus, $\{(-1)^{r_e} r_e \bmod p : e \in E\} \subseteq E$. The set $\{(-1)^{r_e} r_e \bmod p : e \in E\}$ can fail to be equal to E only if $(-1)^{r_e} r_e \equiv (-1)^{r_f} r_f \pmod{p}$ for two distinct elements e and f in E. This implies that $(-1)^{r_e} qe \equiv (-1)^{r_f} qf$, which implies that $e \equiv \pm f$. Since e and f are distinct, $e \equiv -f$ or $e + f \equiv 0$, that is, $p | e + f$. But $0 < e + f < 2p$, and $e + f$ is even while p is odd, so this is impossible. This contradiction implies the claim.

Now

$$q^{(p-1)/2} \prod_{e \in E} e = \prod_{e \in E} qe \equiv \prod_{e \in E} r_e$$

and

$$\prod_{e \in E} e \equiv \prod_{e \in E} (-1)^{r_e} r_e = (-1)^{\Sigma r} \prod_{e \in E} r_e,$$

where $\sum r = \sum_{e \in E} r_e$. Hence,

$$q^{(p-1)/2} (-1)^{\Sigma r} \prod_{e \in E} r_e \equiv \prod_{e \in E} r_e,$$

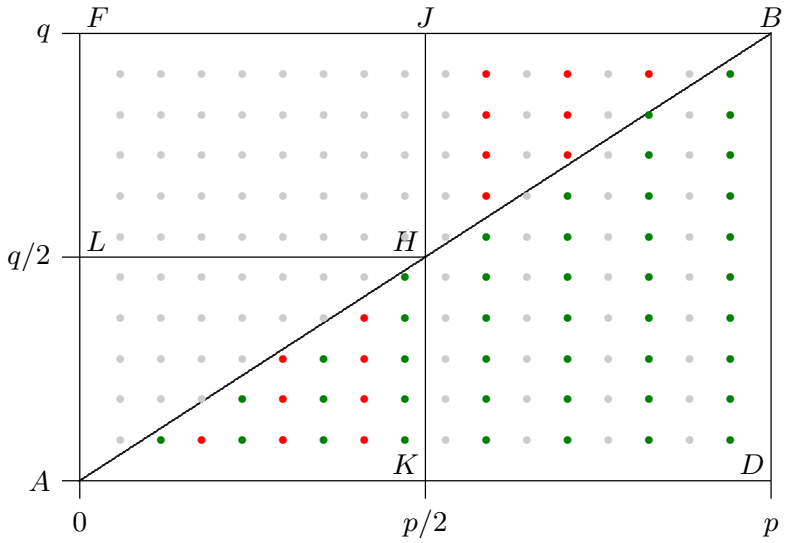
which implies that $q^{(p-1)/2} \equiv (-1)^{\Sigma r}$. By Euler's Criterion, $\left(\frac{q}{p}\right) = (-1)^{\Sigma r}$. Since the value of $(-1)^{\Sigma r}$ is determined by the parity of $\sum_{e \in E} r_e$, we can replace $\sum_{e \in E} r_e$ by any number of the same parity without changing $(-1)^{\Sigma r}$.

Note that

$$\sum_{e \in E} qe = \sum_{e \in E} (p \lfloor qe/p \rfloor + r_e) = p \sum_{e \in E} \lfloor qe/p \rfloor + \sum_{e \in E} r_e,$$

so $\sum_{e \in E} \lfloor qe/p \rfloor$ has the same parity as $\sum_{e \in E} r_e$. (Recall that the floor function $\lfloor \, \rfloor$ means to round down to the nearest integer.)

In what follows we refer to this diagram of a portion of the first quadrant:



The **integer lattice points** are the points in the plane with two integer coordinates; we are interested in the integer lattice points inside rectangle $AFBD$, *not* including those on the edges. In the diagram, we show the integer lattice points using $p = 17$ and $q = 11$.

Because p and q are prime, there are no integer lattice points on the line AB. For any $e \in E$, the point $(e, qe/p)$ is on the line AB, so the number of integer lattice points inside

triangle ABD with abscissa e is $\lfloor qe/p \rfloor$, and $\sum_{e\in E}\lfloor qe/p\rfloor$ is the number of integer lattice points in ABD with even abscissas; these are shown in green in the diagram.

The number of integer lattice points inside rectangle $ADBF$ with a given integer abscissa is even (namely, $q-1$), so the number of these points above line AB has the same parity as the number below AB. Suppose $e \in E$ and $e > p/2$. The number of integer lattice points with abscissa e above line AB is the same as the number of integer lattice points with abscissa $p-e$ below AB; these points are shown in red. Since $p-e$ is odd,

$$\begin{aligned}
\sum_{e\in E}\left\lfloor \frac{qe}{p}\right\rfloor &= \sum_{e<p/2}\left\lfloor \frac{qe}{p}\right\rfloor + \sum_{e>p/2}\left\lfloor \frac{qe}{p}\right\rfloor \\
&\equiv \sum_{e<p/2}\left\lfloor \frac{qe}{p}\right\rfloor + (\text{the number of points with even abscissa in } HJB) \\
&= \sum_{e<p/2}\left\lfloor \frac{qe}{p}\right\rfloor + (\text{the number of points with odd abscissa in } AKH) \\
&= (\text{the number of lattice points in } AKH) \\
&\doteq \mu,
\end{aligned}$$

where the congruence on the second line is (mod 2). (The symbol $\doteq$ indicates that we are defining a new variable μ equal to the number of lattice points in AKH.) So μ has the same parity as $\sum_{e\in E} r_e$, and $\left(\frac{q}{p}\right) = (-1)^\mu$. By precisely the same argument, $\left(\frac{p}{q}\right) = (-1)^\nu$, where ν is the number of integer lattice points in ALH. Since there are $((p-1)/2)((q-1)/2)$ integer lattice points in rectangle $AKHL$, we have

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\mu+\nu} = (-1)^{((p-1)/2)((q-1)/2)},$$

as promised. ∎

Ferdinand Gotthold Max Eisenstein. Eisenstein (1823-1852) was born to parents of limited means and remained near poverty throughout his life. He had five younger siblings, all of whom died in childhood—most of meningitis, which also afflicted Eisenstein. He suffered from poor health and depression for most of his life.

Eisenstein first became interested in mathematics when he was six, thanks to a family acquaintance. In his autobiography, Eisenstein wrote, "As a boy of six I could understand the proof of a mathematical theorem more readily than that meat had to be cut with one's

knife, not one's fork." He also had a lifelong interest in music—he played the piano and composed.

Eisenstein had some excellent and encouraging teachers in mathematics, and began reading the work of Euler, Lagrange and Gauss at an early age. In 1843 he passed his secondary school examinations, though he already knew far more mathematics than the standard secondary fare. He enrolled at the University of Berlin and submitted his first paper in January of 1844. In that year, volumes 27 and 28 of Crelle's mathematical journal contained *twenty-five* works by Eisenstein, making him an overnight sensation in mathematical circles. Gauss was very impressed by Eisenstein's early work, and wrote the preface for an 1847 collection of work by Eisenstein.

Through Crelle, Eisenstein met Alexander von Humboldt, who became his mentor, champion and financial lifeline. Humboldt secured a series of small grants for Eisenstein, and sometimes contributed his own funds to help Eisenstein through times between grants.

Eisenstein was minimally involved in the political unrest of 1848. He was arrested and detained overnight, suffering severe mistreatment that hurt his already poor health. The incident also made it even more difficult for him to find financial support; Humboldt was just barely able to find some funding for him. Eisenstein's health deteriorated and his depression increased, so that he was often unable to deliver his lectures, but he continued to publish papers.

In 1851 Eisenstein was elected to the Göttingen Society, and in 1852 to the Berlin Academy. In July of 1852, his health declined precipitously when he suffered a hemorrhage. Humboldt raised enough money to send him to recuperate in Italy for a year, but it came too late. Eisenstein died in October of tuberculosis.

Our exposition of Eisenstein's proof is taken from *Eisenstein's Misunderstood Geometric Proof of the Quadratic Reciprocity Theorem*, by Reinhard Laubenbacher and David Pengelley, in The College Mathematics Journal, volume 25, number 1, January 1994. Biographical information is from the same paper, and from the article on Eisenstein, by Kurt-R. Biermann, in Biographical Dictionary of Mathematicians, New York: Charles Scribner's Sons, 1991.

Exercises 3.12.

1. Verify the quadratic reciprocity theorem directly for the following pairs of primes. That is, compute $\left(\frac{q}{p}\right)$ and $\left(\frac{p}{q}\right)$ directly by determining whether or not each is a quadratic residue modulo the other, and then check that the theorem is satisfied.

a) 5, 11

b) 3, 19

2. Prove that $x^2 \equiv y^2 \pmod{p}$ implies that $y \equiv \pm x \pmod{p}$, for prime p. Hint: look at the proof of Theorem 3.10.1.

3. Prove Wilson's Theorem 3.10.3 from Theorem 3.12.1.

4. Explain why Euler's Criterion is implied by the last congruence in the proof of Corollary 3.12.2.

5. Prove that there are no integer lattice points on AB.

6. Prove that the number of integer lattice points with abscissa e above line AB is the same as the number of integer lattice points with abscissa $p - e$ below AB.

7. The Quadratic Reciprocity Theorem can be restated in a different, perhaps more appealing, way:

Suppose p and q are distinct odd primes. Then p and q are each quadratic residues of the other, or are each quadratic non-residues of the other, unless both $(p-1)/2$ and $(q-1)/2$ are odd.

Prove this version of the theorem, using theorem 3.12.3.

4

Functions

4.1 Definition and Examples

You have certainly dealt with functions before, primarily in calculus, where you studied functions from $\mathbb{R}$ to $\mathbb{R}$ or from $\mathbb{R}^2$ to $\mathbb{R}$. Perhaps you have encountered functions in a more abstract setting as well; this is our focus. In the last few sections of the chapter, we use functions to study some interesting topics in set theory.

By a **function** from a set A to a set B we mean an *assignment* or *rule* f such that for every $a \in A$ there is a unique $b \in B$ such that $f(a) = b$. The set A is called the **domain** of f and the set B is called the **codomain**. We say two functions f and g are **equal** if they have the same domain and the same codomain, and if for every a in the domain, $f(a) = g(a)$.

(In the interest of full disclosure of dirty tricks, we should mention that the last paragraph is not really a definition at all! The problem is that the words "assignment" and "rule" are synonyms for "function." This problem can be "resolved" by defining functions in terms of sets, but we don't really have a satisfactory definition of "set" either. For now, all that is needed is an intuitive understanding of the concept and a way of showing two functions are equal.)

We often write $f\colon A \to B$ to indicate that f is a function from A to B. Sometimes the word "map" or "mapping" is used instead of "function." If $f\colon A \to B$ and $f(a) = b$, we say b is the **image of** a **under** f, and a is a **preimage of** b **under** f. When the function is clear from context, the phrase 'under f' may be dropped.

EXAMPLE 4.1.1 You are familiar with many functions $f:\mathbb{R} \to \mathbb{R}$: Polynomial functions, trigonometric functions, exponential functions, and so on. Often you have dealt with functions with codomain $\mathbb{R}$ whose domain is some subset of $\mathbb{R}$. For example, $f(x) = \sqrt{x}$ has domain $[0, \infty)$ and $f(x) = 1/x$ has domain $\{x \in \mathbb{R} : x \neq 0\}$. It is easy to see that a subset of the plane is the graph of a function $f:\mathbb{R} \to \mathbb{R}$ if and only if every vertical line intersects it at exactly one point. If this point is (a, b), then $f(a) = b$. □

EXAMPLE 4.1.2 Functions on finite sets can be defined by listing all the assignments. If $A = \{1, 2, 3, 4\}$ and $B = \{r, s, t, u, v\}$ then "$f(1) = t, f(2) = s, f(3) = u, f(4) = t$" defines a function from A to B. The assignment can be done quite arbitrarily, without recourse to any particular formula. □

EXAMPLE 4.1.3 The following are not functions from $A = \{1, 2, 3, 4, 5\}$ to $B = \{r, s, t, u\}$:

$$\begin{array}{ll} f(1) = t & g(1) = u \\ f(2) = s & g(2) = r \\ f(3) = r & g(4) = s \\ f(3) = u & g(5) = t \\ f(4) = u & \\ f(5) = r & \end{array}$$

The problem is that f maps 3 to two values and g doesn't map 3 to any values. When listing the assignments for a function the elements of the domain must appear exactly once. (Elements of the codomain may appear more than once or not at all. In example 4.1.2, the element t of the codomain has two preimages and r and v have none. We will discuss this situation at length in later sections.) □

EXAMPLE 4.1.4 If A and B are non-empty sets and b_0 is a fixed element of B, we can define a **constant** function $f:A \to B$ by the formula $f(a) = b_0$ for all $a \in A$. There are as many constant functions from A to B as there are elements of B. □

EXAMPLE 4.1.5 For a set A we define the **identity** function $i_A:A \to A$ by the rule $i_A(a) = a$ for all $a \in A$. In other words, the identity function maps every element to itself. Though this seems like a rather trivial concept, it is useful and important. Identity functions behave in much the same way that 0 does with respect to addition or 1 does with respect to multiplication. □

EXAMPLE 4.1.6 If $A \subseteq B$, define the **inclusion** function $f:A \to B$ by $f(a) = a$ for every $a \in A$. This is very similar to i_A; the only difference is the codomain. □

DEFINITION 4.1.7 If $f: A \to B$ and $g: B \to C$ are functions, define $g \circ f: A \to C$ by the rule $(g \circ f)(a) = g(f(a))$ for all $a \in A$. This is called the **composition** of the two functions. Observe that f is the first function that is applied to an element a though it is listed on the right. This violation of the usual left-to-right convention sometimes causes confusion. □

EXAMPLE 4.1.8 If $f: \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is given by $f(x) = \sqrt{x}$ and $g: \mathbb{R} \to \mathbb{R}$ is given by $g(x) = \sin x$, then $g \circ f: \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is given by $(g \circ f)(x) = \sin \sqrt{x}$. Note that $(f \circ g)(x) = \sqrt{\sin x}$ makes sense only for those x such that $\sin x \geq 0$. In general, $f \circ g$ and $g \circ f$ are not necessarily equal, and (as in this case) they need not be defined at the same points. □

EXAMPLE 4.1.9 If $A = \{1, 2, 3, 4\}$, $B = \{r, s, t, u\}$, $C = \{\$, \%, \#, \&\}$, and

$$\begin{array}{llll} f(1) & = u & g(r) & = \% \\ f(2) & = r & g(s) & = \# \\ f(3) & = s & g(t) & = \$ \\ f(4) & = u & g(u) & = \$ \end{array}$$

then

$$\begin{aligned} (g \circ f)(1) &= \$ \\ (g \circ f)(2) &= \% \\ (g \circ f)(3) &= \# \\ (g \circ f)(4) &= \$ \end{aligned}$$

□

EXAMPLE 4.1.10 If $A \subseteq B$, $f: A \to B$ is the inclusion function (example 4.1.6) and $g: B \to C$ is a function, then $g \circ f: A \to C$ is called the **restriction** of g to A and is usually written $g|_A$. For all $a \in A$,

$$g|_A(a) = g(f(a)) = g(a),$$

so $g|_A$ is just the same function as g with a smaller domain. □

The following is an easy but important observation:

THEOREM 4.1.11 If $f: A \to B$ then $f \circ i_A = f = i_B \circ f$.

Proof. All three functions have domain A and codomain B. For every $a \in A$,

$$(f \circ i_A)(a) = f(i_A(a)) = f(a) = i_B(f(a)) = (i_B \circ f)(a).$$

■

A similar argument shows that whenever it is defined, composition of functions is associative, i.e., $(f \circ g) \circ h = f \circ (g \circ h)$ (see exercise 7).

Exercises 4.1.

1. Decide if the following assignments define functions from $A = \{1, 2, 3, 4\}$ to $B = \{r, s, t, u, v\}$.

$f(1) = s$	$g(1) = t$	$h(1) = v$
$f(2) = t$	$g(2) = r$	$h(2) = u$
$f(4) = u$	$g(3) = s$	$h(3) = t$
	$g(4) = r$	$h(2) = s$
		$h(4) = r$

2. Let $f\colon \{s, t, u, v, w, x\} \to \{1, 2, 3, 4, 5\}$ and $g\colon \{1, 2, 3, 4, 5\} \to \{m, n, o, p\}$ be given by:

$f(s) = 2$	$g(1) = m$
$f(t) = 1$	$g(2) = n$
$f(u) = 4$	$g(3) = p$
$f(v) = 2$	$g(4) = o$
$f(w) = 1$	$g(5) = m$
$f(x) = 2$	

Find the following:

a) $h = g \circ f$

b) The image of u under f

c) The image of 2 under g

d) The image of v under h

e) The preimage(s) of p under g

f) The preimage(s) of 1 under f

g) The preimage(s) of n under h

h) The preimage(s) of 5 under f

3. Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is given by $f(x) = \cos x$ and $g\colon \mathbb{R} \to \mathbb{R}$ is given by $g(x) = x^2$. Find the following:

a) $h = g \circ f$

b) The image of 4π under f

c) The image of $-\sqrt{2}$ under g

d) The image of $\pi/4$ under h

e) The preimage(s) of $\sqrt{3}/2$ under f

f) The preimage(s) of $9/25$ under g

g) The preimage(s) of 1 under h

h) The preimage(s) of 2 under f

4. Suppose f and g are both functions from A to A. If $f \circ f = g \circ g$, does it follow that $f = g$?

5. Suppose A and B are finite non-empty sets with m and n elements respectively. How many functions are there from A to B?

6. Suppose f and g are two functions from A to B. If $A = X \cup Y$, prove $f = g$ iff $f|_X = g|_X$ and $f|_Y = g|_Y$.

7. Suppose $f\colon C \to D$, $g\colon B \to C$ and $h\colon A \to B$ are functions. Prove $(f \circ g) \circ h = f \circ (g \circ h)$.

4.2 Induced Set Functions

Sets and functions are intimately related, as we will see throughout the chapter. Here we begin to explore some basic connections. Suppose $f\colon A \to B$ is a function. If $X \subseteq A$, define

$$f(X) = \{b \in B : \exists\, a \in X (b = f(a))\} \subseteq B,$$

called the **image** of X. If $Y \subseteq B$, define

$$f^{-1}(Y) = \{a \in A : f(a) \in Y\} \subseteq A,$$

called the **preimage** of Y. There is real opportunity for confusion here: the letter f is being used in two different, though related, ways. We can apply f to *elements* of A to get *elements* of B, or we can apply it to *subsets* of A to get *subsets* of B. Similarly, we can talk about the images or preimages of either elements or subsets. Context should always make it clear what is meant, but you should be aware of the problem.

EXAMPLE 4.2.1

Suppose $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{r, s, t, u, v, w\}$ and

$$\begin{array}{lll} f(1) = r & f(3) = v & f(5) = r \\ f(2) = s & f(4) = t & f(6) = v \end{array}$$

Then

$$f(\{1, 3, 5\}) = \{r, v\},$$
$$f(\{4, 5, 6\}) = \{t, r, v\},$$

and

$$f^{-1}(\{r, t, u\}) = f^{-1}(\{r, t\}) = \{1, 4, 5\},$$
$$f^{-1}(\{u, w\}) = \emptyset.$$

□

EXAMPLE 4.2.2 Suppose $f: \mathbb{R} \to \mathbb{R}$ is given by $f(x) = x^2$. Then

$$f([2, 3]) = [4, 9],$$
$$f((-2, 1]) = [0, 4),$$
$$f(\{1, 2, 3\}) = \{1, 4, 9\},$$

and

$$f^{-1}([0, 1]) = [-1, 1],$$
$$f^{-1}([-1, 0]) = \{0\},$$
$$f^{-1}((-\infty, 0)) = \emptyset.$$

□

By the **range** (or **image**) of a function $f: A \to B$, we mean

$$f(A) = \{b \in B : \exists a \in A \, (b = f(a))\}.$$

The range may be considerably smaller than the codomain.

EXAMPLE 4.2.3 The range of the function in example 4.2.1 is $\{r, s, v, t\}$, which is a proper subset of the codomain. □

EXAMPLE 4.2.4 The range of $\sin: \mathbb{R} \to \mathbb{R}$ is $[-1, 1]$. The range of $f: \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is $[0, \infty)$. □

The next two theorems show how induced set functions behave with respect to intersection and union.

THEOREM 4.2.5 Suppose $f: A \to B$ is a function and Y and Z are subsets of B. Then

a) $f^{-1}(Y \cup Z) = f^{-1}(Y) \cup f^{-1}(Z)$,

b) $f^{-1}(Y \cap Z) = f^{-1}(Y) \cap f^{-1}(Z)$.

Proof. We prove part (b) and leave part (a) as an exercise. If $a \in A$, then $a \in f^{-1}(Y \cap Z)$ if and only if $f(a)$ is in $Y \cap Z$. This is true if and only if $f(a) \in Y$ and $f(a) \in Z$. This, in turn, is equivalent to $a \in f^{-1}(Y)$ and $a \in f^{-1}(Z)$. Finally, this is true if and only if $a \in f^{-1}(Y) \cap f^{-1}(Z)$. ■

THEOREM 4.2.6 Suppose $f: A \to B$ is a function and W and X are subsets of A. Then

a) $f(W \cup X) = f(W) \cup f(X)$,

b) $f(W \cap X) \subseteq f(W) \cap f(X)$.

Proof. We'll do part (b). If $b \in B$ is in $f(W \cap X)$, then $b = f(a)$ for some $a \in W \cap X$. Since $a \in W \cap X$, a is in both W and X. Therefore, $b = f(a)$ is in both $f(W)$ and $f(X)$, that is, $b \in f(W) \cap f(X)$. ■

It is perhaps surprising to compare these two theorems and observe that of the two induced set functions, it is f^{-1} that is "better behaved" with respect to the usual set operations.

Exercises 4.2.

In the first two exercises, use the function $f: \{1, 2, 3, 4, 5, 6, 7\} \to \{a, b, c, d, e, f, g, h\}$ given by:

$$\begin{array}{lll} f(1) = d & f(4) = a & f(6) = e \\ f(2) = e & f(5) = b & f(7) = f \\ f(3) = f & & \end{array}$$

1. Find the following:

a) $f(\{2, 4, 6\})$

b) $f(\emptyset)$

c) $f^{-1}(\{d, e, h\})$

d) $f^{-1}(\{a, b, f, c\})$

2. Use the function f given above.

a) If $Y = \{a, b, c, e, f\}$ and $Z = \{a, b, e, g, h\}$, verify the statements in Theorem 4.2.5.

b) If $W = \{1, 2, 3, 4\}$ and $X = \{2, 4, 6, 7\}$, verify the statements in Theorem 4.2.6.

3. Suppose $f: \mathbb{R} \to \mathbb{R}$ is given by $f(x) = |x - 1|$. Find the following:

a) $f([-1, 1])$

b) $f(\{-4, -2, 0, 1, 5\})$

c) $f^{-1}((0, 2))$

d) $f^{-1}(\{-2, 0, 4, 5\})$

4. Suppose $f: \mathbb{R} \to \mathbb{R}$ is given by $f(x) = x^2$.

a) If $Y = (1, \infty)$ and $Z = (-\infty, 4)$, verify the statements in Theorem 4.2.5.

b) If $W = [-3, 2]$, and $X = (0, 4]$, verify the statements in Theorem 4.2.6.

5. Prove 4.2.5(a).

6. Prove 4.2.6(a).

In the next two exercises suppose $f: A \to B$ is a function and $\{X_i\}_{i \in I}$ is a family of subsets of A.

7. Prove $f(\bigcup_{i \in I} X_i) = \bigcup_{i \in I} f(X_i)$.

8. Prove $f(\bigcap_{i \in I} X_i) \subseteq \bigcap_{i \in I} f(X_i)$.

In the next two exercises suppose $\{Y_i\}_{i \in I}$ is an indexed family of subsets of B and $f: A \to B$ is a function.

9. Prove $f^{-1}(\bigcup_{i \in I} Y_i) = \bigcup_{i \in I} f^{-1}(Y_i)$.

10. Prove $f^{-1}(\bigcap_{i \in I} Y_i) = \bigcap_{i \in I} f^{-1}(Y_i)$.

4.3 Injections and Surjections

Two simple properties that functions may have turn out to be exceptionally useful. If the codomain of a function is also its range, then the function is **onto** or **surjective**. If a function does not map two different elements in the domain to the same element in the range, it is **one-to-one** or **injective**. In this section, we define these concepts "officially" in terms of preimages, and explore some easy examples and consequences.

DEFINITION 4.3.1 A function $f: A \to B$ is injective if each $b \in B$ has at most one preimage in A, that is, there is at most one $a \in A$ such that $f(a) = b$. □

EXAMPLE 4.3.2 Suppose $A = \{1, 2, 3\}$ and $B = \{r, s, t, u, v\}$ and

$$\begin{aligned} f(1) &= s & g(1) &= r \\ f(2) &= t & g(2) &= t \\ f(3) &= r & g(3) &= r. \end{aligned}$$

Here f is injective since r, s, t have one preimage and u, v have no preimages. On the other hand, g fails to be injective, since r has more than one preimage. □

EXAMPLE 4.3.3 Define $f, g : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$, $g(x) = 2^x$. Function f fails to be injective because any positive number has two preimages (its positive and negative square roots). On the other hand, g is injective, since if $b \in \mathbb{R}$, then $g(x) = b$ has at most one solution (if $b > 0$ it has one solution, $\log_2 b$, and if $b \leq 0$ it has no solutions). □

EXAMPLE 4.3.4 If $A \subseteq B$, then the inclusion map from A to B is injective. □

An injective function is called an **injection**. An injection may also be called a one-to-one (or 1–1) function; some people consider this less formal than "injection".

There is another way to characterize injectivity which is useful for doing proofs. To say that the elements of the codomain have at most one preimage is to say that no two elements of the domain are taken to the same element, as we indicated in the opening paragraph. In other words, $f: A \to B$ is injective if and only if for all $a, a' \in A$, $a \neq a'$ implies $f(a) \neq f(a')$. Taking the contrapositive, f is injective if and only if for all $a, a' \in A$, $f(a) = f(a')$ implies $a = a'$.

THEOREM 4.3.5 If $f: A \to B$ and $g : B \to C$ are injective functions, then $g \circ f: A \to C$ is injective also.

Proof. Suppose $g(f(a)) = g(f(a'))$. Since g is injective, $f(a) = f(a')$. Since f is injective, $a = a'$. Thus, $(g \circ f)(a) = (g \circ f)(a')$ implies $a = a'$, so $(g \circ f)$ is injective. ■

DEFINITION 4.3.6 A function $f: A \to B$ is surjective if each $b \in B$ has at least one preimage, that is, there is at least one $a \in A$ such that $f(a) = b$. □

EXAMPLE 4.3.7 Suppose $A = \{1, 2, 3, 4, 5\}$, $B = \{r, s, t\}$, and

$$\begin{array}{ll} f(1) = s & g(1) = t \\ f(2) = r & g(2) = r \\ f(3) = s & g(3) = r \\ f(4) = t & g(4) = t \\ f(5) = r & g(5) = t \end{array}$$

Under f, the elements r, s, t have 2, 2, and 1 preimages, respectively, so f is surjective. Under g, the element s has no preimages, so g is not surjective. □

EXAMPLE 4.3.8 Define $f, g : \mathbb{R} \to \mathbb{R}$ by $f(x) = 3^x$, $g(x) = x^3$. Since 3^x is always positive, f is not surjective (any $b \leq 0$ has no preimages). On the other hand, for any $b \in \mathbb{R}$ the equation $b = g(x)$ has a solution (namely $x = \sqrt[3]{b}$) so b has a preimage under g. Therefore g is surjective. □

EXAMPLE 4.3.9 Suppose A and B are sets with $A \neq \emptyset$. Then $p : A \times B \to B$ given by $p((a, b)) = b$ is surjective, and is called the **projection onto** B. □

EXAMPLE 4.3.10 For any set A the identity map i_A is both injective and surjective. □

A surjective function is called a **surjection**. A surjection may also be called an onto function; some people consider this less formal than "surjection". To say that a function $f : A \to B$ is a surjection means that every $b \in B$ is in the range of f, that is, the range is the same as the codomain, as we indicated above.

THEOREM 4.3.11 Suppose $f : A \to B$ and $g : B \to C$ are surjective functions. Then $g \circ f : A \to C$ is surjective also.

Proof. Suppose $c \in C$. Since g is surjective, there is a $b \in B$ such that $g(b) = c$. Since f is surjective, there is an $a \in A$, such that $f(a) = b$. Hence $c = g(b) = g(f(a)) = (g \circ f)(a)$, so $g \circ f$ is surjective. ■

Exercises 4.3.

1. Decide if the following functions from $\mathbb{R}$ to $\mathbb{R}$ are injections, surjections, or both.

a) $2x + 1$
b) $1/2^x$
c) $\sin x$
d) $(x + 1)^3$
e) $x^3 - x$
f) $|x|$

2. **a)** Find an example of an injection $f: A \to B$ and a surjection $g: B \to C$ such that $g \circ f$ is neither injective nor surjective.

b) Find an example of a surjection $f: A \to B$ and an injection $g: B \to C$ such that $g \circ f$ is neither injective nor surjective.

3. **a)** Suppose A and B are finite sets and $f: A \to B$ is injective. What conclusion is possible regarding the number of elements in A and B? Justify your answer.

b) If instead of injective, we assume f is surjective, what conclusion is possible? Justify your answer.

4. Suppose A is a finite set. Can we construct a function $f: A \to A$ that is injective, but not surjective? Surjective, but not injective?

5. **a)** Find a function $f: \mathbb{N} \to \mathbb{N}$ that is injective, but not surjective.

b) Find a function $g: \mathbb{N} \to \mathbb{N}$ that is surjective, but not injective.

6. Suppose A and B are non-empty sets with m and n elements respectively, where $m \leq n$. How many injective functions are there from A to B?

7. Find an injection $f: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. (Hint: use prime factorizations.)

8. If $f: A \to B$ is a function, $A = X \cup Y$ and $f|_X$ and $f|_Y$ are both injective, can we conclude that f is injective?

4.4 More Properties of Injections and Surjections

Injections and surjections are 'alike but different,' much as intersection and union are 'alike but different.' This is another example of *duality.*

THEOREM 4.4.1 Suppose $f: A \to B$ and $g: B \to C$ are functions.

a) If $g \circ f$ is injective then f is injective.

b) If $g \circ f$ is surjective then g is surjective.

Proof. We prove part (a), leaving part (b) as an exercise. Suppose $a, a' \in A$ and $f(a) = f(a')$. We wish to prove $a = a'$. We have

$$(g \circ f)(a) = g(f(a)) = g(f(a')) = (g \circ f)(a')$$

and since $g \circ f$ is injective, we conclude $a = a'$, as desired. ∎

The next result shows that injective and surjective functions can be "canceled." As in theorem 4.4.1, the result in the two cases is 'the same, but different.'

THEOREM 4.4.2 Suppose $f_1, f_2: A \to B$, $g: B \to C$, $h_1, h_2: C \to D$ are functions.

a) If g is injective and $g \circ f_1 = g \circ f_2$ then $f_1 = f_2$.

b) If g is surjective and $h_1 \circ g = h_2 \circ g$ then $h_1 = h_2$.

Proof. We prove part (b), leaving part (a) as an exercise. Suppose $c \in C$. We wish to show $h_1(c) = h_2(c)$. By hypothesis g is surjective, so there is a $b \in B$ such that $g(b) = c$. So

$$h_1(c) = h_1(g(b)) = (h_1 \circ g)(b) = (h_2 \circ g)(b) = h_2(g(b)) = h_2(c),$$

as desired. ■

Exercises 4.4.

1. Show by example that if $g \circ f$ is injective, then g need not be injective.
2. Show by example that if $g \circ f$ is surjective, then f need not be surjective.
3. Show by example that a function g that is not injective may not be "cancellable" when it appears on the left, i.e., there may exist $f_1 \neq f_2$ such that $g \circ f_1 = g \circ f_2$.
4. Show by example that a function f that is not surjective may not be "cancellable" when it appears on the right, i.e., there may exist $g_1 \neq g_2$ such that $g_1 \circ f = g_2 \circ f$.
5. Prove 4.4.1(b).
6. Prove 4.4.2(a).
7. Suppose $f: A \to B$ is a surjection and $Y \subseteq B$. Show that $f(f^{-1}(Y)) = Y$.
8. Suppose $f: A \to B$ is injective and W, X are disjoint subsets of A. Prove that $f(W)$ and $f(X)$ are disjoint subsets of B.

4.5 Pseudo-Inverses

Suppose $f: A \to B$ is a function with range R. A function $g: B \to A$ is a **pseudo-inverse** of f if for all $b \in R$, $g(b)$ is a preimage of b.

EXAMPLE 4.5.1 If $A = \{1, 2, 3, 4\}$, $B = \{r, s, t\}$ and

$$f(1) = r \qquad f(2) = t \qquad f(3) = t \qquad f(4) = r$$

then $R = \{r, t\}$ and

$$g(r) = 4 \qquad g(s) = 3 \qquad g(t) = 2$$

is a pseudo-inverse to f; there are others, of course. The important point is that g must map r to either 1 or 4, and t to either 2 or 3. We will usually be interested in the pseudo-inverse when f is injective or surjective. □

THEOREM 4.5.2 If f is injective, any pseudo-inverse is surjective; if f is surjective, any pseudo-inverse is injective.

Proof. Suppose f is injective, and that a is any element of A. Then $f(a)$ is an element of the range of f, which we denote by b. If g is a pseudo-inverse to f, then $g(b)$ must be a preimage of b, but since f is injective, b has only one preimage, namely a. So

$g(f(a)) = g(b) = a$. In other words, $g \circ f = i_A$ and we say g is a **left inverse** of f. By theorem 4.4.1, g is surjective.

Suppose f is surjective. In this case, $R = B$, so for any $b \in B$, $g(b)$ is a preimage of b. This means that $f(g(b)) = b$. In other words, $f \circ g = i_B$. We say that g is a **right inverse** to f when this happens. By theorem 4.4.1, g is injective. ■

EXAMPLE 4.5.3 If $A = \{1, 2, 3, 4, 5\}$, $B = \{r, s, t\}$ and

$$\begin{array}{lll} f(1) = r & f(3) = t & f(5) = s \\ f(2) = t & f(4) = r & \end{array}$$

then

$$g(r) = 4 \qquad g(s) = 5 \qquad g(t) = 2$$

is a pseudo-inverse to f. It is easy to check that $f \circ g = i_B$. □

EXAMPLE 4.5.4 If $A = \{1, 2, 3, 4\}$, $B = \{r, s, t, u, v, w\}$ and

$$f(1) = s \qquad f(2) = v \qquad f(3) = w \qquad f(4) = r$$

then

$$\begin{array}{lll} g(r) = 4 & g(t) = 2 & g(v) = 2 \\ g(s) = 1 & g(u) = 4 & g(w) = 3 \end{array}$$

is a pseudo-inverse to f. It is easy to check that $g \circ f = i_A$. □

Exercises 4.5.

1. Find pseudo-inverses for the following functions:

a) $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{r, s, t, u\}$

$$\begin{array}{lll} f(1) = t & f(3) = u & f(5) = u \\ f(2) = t & f(4) = s & f(6) = s \end{array}$$

b) $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{r, s, t\}$

$$\begin{array}{lll} f(1) = r & f(3) = t & f(5) = s \\ f(2) = s & f(4) = t & f(6) = s \end{array}$$

c) $A = \{1, 2, 3, 4\}$, $B = \{r, s, t, u, v, w\}$

$$\begin{array}{ll} f(1) = t & f(3) = u \\ f(2) = r & f(4) = s \end{array}$$

d) $f: \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$.

e) $f: \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$.

2. Determine whether the pseudo-inverses for the functions listed in problem 1 are right inverses, left inverses, both, or neither.

3. Prove that every function $f: A \to B$ has a pseudo-inverse.

4. Give a proof of Theorem 4.4.2 using pseudo-inverses.

5. How many pseudo-inverses do each of the functions in 1(a,b,c) have?

6. If g is a pseudo-inverse to f, what is $f \circ g \circ f$?

7. If A has 4 elements and B has 3 elements, what is the least number of pseudo-inverses that a function $f: A \to B$ might have? What is the greatest number?

4.6 Bijections and Inverse Functions

A function $f: A \to B$ is **bijective** (or f is a **bijection**) if each $b \in B$ has exactly one preimage. Since *"at least one"* + *"at most one"* = *"exactly one"*, f is a bijection if and only if it is both an injection and a surjection. A bijection is also called a **one-to-one correspondence**.

EXAMPLE 4.6.1 If $A = \{1, 2, 3, 4\}$ and $B = \{r, s, t, u\}$, then

$$f(1) = u \qquad f(3) = t$$
$$f(2) = r \qquad f(4) = s$$

is a bijection. □

EXAMPLE 4.6.2 The functions $f: \mathbb{R} \to \mathbb{R}$ and $g: \mathbb{R} \to \mathbb{R}^+$ (where $\mathbb{R}^+$ denotes the positive real numbers) given by $f(x) = x^5$ and $g(x) = 5^x$ are bijections. □

EXAMPLE 4.6.3 For any set A, the identity function i_A is a bijection. □

DEFINITION 4.6.4 If $f: A \to B$ and $g: B \to A$ are functions, we say g is an **inverse** to f (and f is an inverse to g) if and only if $f \circ g = i_B$ and $g \circ f = i_A$. □

EXAMPLE 4.6.5 If f is the function from example 4.6.1 and

$$g(r) = 2 \qquad g(t) = 3$$
$$g(s) = 4 \qquad g(u) = 1$$

then f and g are inverses. For example, $f(g(r)) = f(2) = r$ and $g(f(3)) = g(t) = 3$. □

EXAMPLE 4.6.6 An inverse to x^5 is $\sqrt[5]{x}$:

$$(\sqrt[5]{x})^5 = x, \quad \sqrt[5]{x^5} = x.$$

□

EXAMPLE 4.6.7 If we think of the exponential function e^x as having domain $\mathbb{R}$ and codomain $\mathbb{R}^{>0}$ (the positive real numbers), and $\ln x$ as having domain $\mathbb{R}^{>0}$ and codomain $\mathbb{R}$, then they are inverses: and

$$\ln e^x = x, \quad e^{\ln x} = x.$$

□

EXAMPLE 4.6.8 The identity function $i_A: A \to A$ is its own inverse. □

If you understand these examples, the following should come as no surprise.

THEOREM 4.6.9 A function $f: A \to B$ has an inverse if and only if it is bijective.

Proof. Suppose g is an inverse for f (we are proving the implication $\Rightarrow$). Since $g \circ f = i_A$ is injective, so is f (by 4.4.1(a)). Since $f \circ g = i_B$ is surjective, so is f (by 4.4.1(b)). Therefore f is injective and surjective, that is, bijective.

Conversely, suppose f is bijective. Let $g: B \to A$ be a pseudo-inverse to f. From the proof of theorem 4.5.2, we know that since f is surjective, $f \circ g = i_B$, and since f is injective, $g \circ f = i_A$. ■

We have talked about "an" inverse of f, but really there is only one.

THEOREM 4.6.10 If $f: A \to B$ has an inverse function then the inverse is unique.

Proof. Suppose g_1 and g_2 are both inverses to f. Then

$$g_1 = g_1 \circ i_B = g_1 \circ (f \circ g_2) = (g_1 \circ f) \circ g_2 = i_A \circ g_2 = g_2,$$

proving the theorem. (See exercise 7 in section 4.1.) ■

Because of theorem 4.6.10, we can talk about "the" inverse of f, assuming it has one; we write f^{-1} for the inverse of f. Note well that this extends the meaning of "f^{-1}", in a potentially confusing way. No matter what function f we are given, the induced set function f^{-1} is defined, but the inverse function f^{-1} is defined only if f is bijective. In other words, f^{-1} is always defined for *subsets* of the codomain, but it is defined for *elements* of the codomain only if f is a bijection.

We close with a pair of easy observations:

THEOREM 4.6.11

a) The composition of two bijections is a bijection.

b) The inverse of a bijection is a bijection.

Proof. Part (a) follows from theorems 4.3.5 and 4.3.11. For part (b), if $f: A \to B$ is a bijection, then since f^{-1} has an inverse function (namely f), f^{-1} is a bijection. ■

Exercises 4.6.

1. Find an example of functions $f: A \to B$ and $g: B \to A$ such that $f \circ g = i_B$, but f and g are not inverse functions.
2. Suppose $[a]$ is a fixed element of $\mathbb{Z}_n$. Define $A_{[a]}: \mathbb{Z}_n \to \mathbb{Z}_n$ by $A_{[a]}([x]) = [a] + [x]$. Show this is a bijection by finding an inverse to $A_{[a]}$.
3. Suppose $[u]$ is a fixed element of $\mathbb{U}_n$. Define $M_{[u]}: \mathbb{Z}_n \to \mathbb{Z}_n$ by $M_{[u]}([x]) = [u] \cdot [x]$. Show this is a bijection by finding an inverse to $M_{[u]}$.
4. Show that for any m, b in $\mathbb{R}$ with $m \neq 0$, the function $L(x) = mx + b$ is a bijection, by finding an inverse.
5. Suppose $f: A \to A$ is a function and $f \circ f$ is bijective. Is f necessarily bijective?
6. Show there is a bijection $f: \mathbb{N} \to \mathbb{Z}$. (Hint: define f separately on the odd and even positive integers.)
7. If $f: A \to B$ and $g: B \to C$ are bijections, prove $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.
8. Suppose $f: A \to B$ is an injection and $X \subseteq A$. Prove $f^{-1}(f(X)) = X$.

4.7 Cardinality and Countability

Here is a seemingly innocuous question: When are two sets A and B the same size? Why, when they have the same number of elements, of course. This is a good answer, except that it turns out to be not at all clear what "same number of elements" actually means when A and B are infinite. Do $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ all have different numbers of elements, or are some of them the same size?

This question really has no answer unless we agree on what "the same size" means. Here is one way (the standard way) to define it: We say the sets A and B have the same size or **cardinality** if there is a bijection $f: A \to B$. If this is the case we write $A \approx B$.

EXAMPLE 4.7.1 If A and B are finite, then $A \approx B$ if and only if A and B have the same number of elements. □

This example shows that the definition of "same size" extends the usual meaning for finite sets, something that we should require of any reasonable definition.

We say a set A is **countably infinite** if $\mathbb{N} \approx A$, that is, A has the same cardinality as the natural numbers. We say A is **countable** if it is finite or countably infinite.

EXAMPLE 4.7.2 The set E of positive even integers is countably infinite: Let $f: \mathbb{N} \to E$ be $f(n) = 2n$. □

EXAMPLE 4.7.3 The set S of positive integers that are perfect squares is countably infinite: Let $f: \mathbb{N} \to S$ be $f(n) = n^2$. □

In the last two examples, E and S are proper subsets of $\mathbb{N}$, but they have the same cardinality. This seeming paradox is in marked contrast to the situation for finite sets. If A is finite and B is a proper subset of A, it is impossible for A and B to have the same number of elements.

If A is a countably infinite set and $f: \mathbb{N} \to A$ is a bijection, then

$$A = \{f(1), f(2), f(3), \ldots\}.$$

In other words, a set is countably infinite if and only if it can be arranged in an infinite sequence.

EXAMPLE 4.7.4 The set $\mathbb{Z}$ of *all* integers is countably infinite: Observe that we can arrange $\mathbb{Z}$ in a sequence in the following way:

$$0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots$$

This corresponds to the bijection $f: \mathbb{N} \to \mathbb{Z}$ defined by

$$f(n) = \begin{cases} n/2, & \text{if } n \text{ is even;} \\ -(n-1)/2, & \text{if } n \text{ is odd.} \end{cases}$$

□

EXAMPLE 4.7.5 The set $\mathbb{Q}^+$ of positive rational numbers is countably infinite: The idea is to define a bijection $g: \mathbb{N} \to \mathbb{Q}^+$ one prime at a time. The positive integer powers of, say, 2 can be paired up with the non-zero integer powers of 2, that is,

$$\begin{array}{cccccccc} 2, & 4, & 8, & 16, & \ldots & 2^k, & \ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \\ 2, & 1/2, & 4, & 1/4, & \ldots & 2^{f(k+1)}, & \ldots \end{array}$$

where f is the bijection between the positive integers and the entire set of integers in example 4.7.4. We do this for every prime in the same way:

$$\begin{array}{cccccccc} p, & p^2, & p^3, & p^4, & \ldots & p^k, & \ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \\ p, & 1/p, & p^2, & 1/p^2, & \ldots & p^{f(k+1)}, & \ldots \end{array}$$

Call this function g, $g(p^k) = p^{f(k+1)}$. Then we extend this to products of prime powers. For example,

$$\begin{aligned} g(3^4 5^5) &= g(3^4)g(5^5) = 5^3/3^2; \\ g(7^{10}11^4 13^7 17) &= (13^4 17)/(7^5 11^2); \\ g(2^8 3^5 5^4 11^2 7^3) &= (3^3 7^2)/(2^4 5^2 11). \end{aligned}$$

In general, then, let $g(1) = 1$ and

$$g(p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}) = p_1^{f(e_1+1)} p_2^{f(e_2+1)} \cdots p_k^{f(e_k+1)} \tag{4.7.1}$$

That g is a bijection is a consequence of the fact that any rational number can be uniquely expressed as a/b, where a and b are positive integers that are relatively prime (so their prime factorizations involve disjoint sets of primes). □

Here are some simple but important properties of cardinality:

THEOREM 4.7.6 Suppose A, B and C are sets. Then

a) $A \approx A$,

b) $A \approx B$ implies $B \approx A$,

c) $A \approx B$ and $B \approx C$ implies $A \approx C$.

Proof. Since $i_A : A \to A$ is a bijection, part (a) follows. If $f : A \to B$ is a bijection, then by theorem 4.6.11, $f^{-1} : B \to A$ is a bijection, so part (b) is true. Similarly, part (c) follows from the fact that the composition of bijections is a bijection. ■

Exercises 4.7.

1. Show that the following sets are countably infinite:

a) The set of multiples of 3.

b) $\{2^q : q \in \mathbb{Q}^+\}$

c) $\{x \in \mathbb{R} : x > 0 \wedge x^2 \in \mathbb{Q}\}$.

2. **a)** Using the bijection of example 4.7.4, find
(i) $f(14)$, (ii) $f(17)$, (iii) $f^{-1}(5)$, (iv) $f^{-1}(-7)$

b) Using the bijection of example 4.7.5, find
(i) $g(72)$, (ii) $g^{-1}(5/18)$, (iii) $g(3^3 5^2 7^4 13^7)$, (iv) $g^{-1}(2^3 7^4 5^{-2} 13^{-5})$

3. Show that the following sets of real numbers have the same cardinality:

a) $(0, 1)$, $(1, \infty)$

b) $(1, \infty)$, $(0, \infty)$.

c) $(0, \infty)$, $\mathbb{R}$

d) $(0, 1)$, $\mathbb{R}$

4. Show that $\mathbb{Q}$ is countably infinite. (Hint: you can arrange $\mathbb{Q}^+$ in a sequence; use this to arrange $\mathbb{Q}$ into a sequence.)
5. Suppose B is a countably infinite set and S is a subset of B. Explain why S is also countable. Suppose $f: A \to B$ is an injection. Explain why A is countable.
6. Show that $\mathbb{N} \times \mathbb{N}$ is countably infinite. (Hint: map $(n, m) \in \mathbb{N} \times \mathbb{N}$ to $2^{n-1}(2m-1)$.)
7. Use problem 6 and induction to prove that $\mathbb{N}^k = \mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}$ is countably infinite.
8. Prove that $\mathbb{Z}^k = \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$ is countably infinite.
9. Any positive rational number other than 1 can be written as $p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}$ in exactly one way, where the e_i are non-zero integers. Write a simple expression (analogous to equation (4.7.1)) for $g^{-1}: \mathbb{Q}^+ \to \mathbb{N}$.

4.8 Uncountability of the Reals

We have seen that many infinite sets that might seem to have different sizes are in fact the same size. Are there infinite sets that are not the same size as the integers? The answer is 'yes', in fact, a resounding 'yes'—there are infinite sets of infinitely many different sizes. We'll begin by showing that one particular set, $\mathbb{R}$, is **uncountable**. The technique we use is the famous **diagonalization process** of Georg Cantor.

THEOREM 4.8.1 $\mathbb{N} \not\approx \mathbb{R}$.

Proof. The proof is by contradiction. If $\mathbb{R}$ were countably infinite the reals could be arranged in a sequence, say r_1, r_2, r_3, ... We show that this cannot be a listing of all the reals by finding a real number that is not on the list. Imagine that the fractional parts of these numbers are written in their decimal form in a list. The list might start something like this:

$$
\begin{aligned}
&.\underline{2}3454167\ldots,\\
&.1\underline{5}367843\ldots,\\
&.86\underline{9}54367\ldots,\\
&.199\underline{1}9423\ldots,\\
&.22453\underline{6}665\ldots,\\
&\vdots
\end{aligned}
$$

Let r be the real number with decimal expansion $0.d_1d_2d_3d_4d_5\ldots$, where $d_i = 1$ unless the ith expansion has a 1 in the ith place to the right of the decimal point, in which case $d_i = 5$. (For the list above, the expansion would be $0.11151\ldots$; the 'diagonal' entries are underlined.) This decimal expansion is different than *every* expansion in the list, so r is not on the list. ∎

This proof is actually a bit trickier than it appears, because two different decimal expansions can represent the same real number. See exercise 14.

We want to be able to talk about the "size" of an infinite set, in much the same way that we talk about the size of a finite set (as in, "The set $\{a, b, c, d, e\}$ has size 5."). With every set A we associate a symbol $\overline{A}$, called the **cardinal number** of A, and we say that $\overline{A} = \overline{B}$ if and only if $A \approx B$.

Some cardinal numbers occur so frequently that they have been given special names: $\overline{\mathbb{N}} = \aleph_0$ ("aleph-naught") and $\overline{\mathbb{R}} = c$ (the size of the "continuum"). In this language, we can say that the "size" of $\mathbb{Q}$ or of $\mathbb{Z}$ is $\aleph_0$, and that the size of $(0, 1) \subseteq \mathbb{R}$ is c.

One familiar feature of finite 'sizes' is that they come in a particular order—that is, if two sizes are different then one is bigger than the other. When can we say that one infinite cardinal number is bigger than another? Here is a natural way: If $\overline{A}$ and $\overline{B}$ are cardinal numbers, define $\overline{A} \leq \overline{B}$ to mean that there is an injection $f: A \to B$.

There is a potential, somewhat subtle problem with this definition. We are defining a relationship between 'sizes' by referring to *particular* sets that have those sizes. What if we were to choose different sets, say A' and B', with the same sizes?

LEMMA 4.8.2 Suppose $\overline{A} = \overline{A'}$ and $\overline{B} = \overline{B'}$. There is an injection $f: A \to B$ if and only if there is an injection $f': A' \to B'$.

Proof. Suppose there is an injection $f: A \to B$. There are bijections $\theta: A' \to A$ and $\phi: B \to B'$. So $f' = \phi \circ f \circ \theta$ is an injection from A' to B'. The converse is similar. ∎

The upshot of this lemma is that the definition does not depend upon which particular set is chosen to represent these two cardinal numbers, that is, "$\leq$" is **well-defined**. This ordering of the cardinal numbers has some familiar properties.

THEOREM 4.8.3 Suppose A, B and C are sets. Then

a) $\overline{A} \leq \overline{A}$;

b) if $\overline{A} \leq \overline{B}$ and $\overline{B} \leq \overline{C}$, then $\overline{A} \leq \overline{C}$.

Proof. For part (a), use the identity map. For part (b), if $f: A \to B$ and $g: B \to C$ are injections, then $g \circ f: A \to C$ is an injection, so $\overline{A} \leq \overline{C}$. ∎

Exercises 4.8.

1. Show that $\mathbb{R}^2$ is not countable.

2. Show that the set S of all infinite sequences of 0's and 1's is not countable (the elements of S are infinitely long strings of the form "00110101110...").

3. Suppose A and B are disjoint countably infinite sets. Show that $A \cup B$ is countably infinite. (Think about arranging things into sequences.)

4. Suppose A is finite and B is countably infinite. Show that $A \cup B$ is countably infinite.
5. Suppose A and B are countably infinite sets. Show that $A \cup B$ is countably infinite.
6. Use exercise 5 and induction to prove that the union of any finite collection of countably infinite sets is countably infinite.
7. Suppose that $\{A_i \mid i \in \mathbb{N}\}$ is a set of non-empty countable sets. Prove that $\bigcup_{i\in\mathbb{N}} A_i$ is countable.
8. Show that the set of all polynomials with coefficients in $\mathbb{Z}$ is countably infinite.
9. The set of all (complex) roots of all polynomials with coefficients in $\mathbb{Z}$ is the **algebraic numbers**. Show that the algebraic numbers are countable. You may use the face that every polynomial has a finite number of roots.
10. Let $\mathcal{I}$ be the set of irrational numbers (i.e., $\{x \in \mathbb{R} : x \notin \mathbb{Q}\}$). Show that $\mathcal{I}$ is not countable. (Use exercise 5.)
11. Suppose A and B are non-empty sets. Show that $\overline{A} \leq \overline{B}$ if and only if there is a surjection $g: B \to A$.
12. Suppose A is a countable set and $f: A \to B$ is a surjective function. Show that B is also countable. (Hint: use exercise 5 of section 4.7.)
13. Use decimal expansions to construct an injection from (0,1) to the irrationals (remember that a number is rational if and only if its decimal terminates or repeats).
14. In the proof of theorem 4.8.1, we constructed a decimal expansion that was not on a given list of decimal expansions. This does not by itself imply that the real number represented by the constructed expansion is not the same as a real number represented by an expansion on the list, because some real numbers have more than one decimal expansion. Explain which real numbers have more than one decimal expansion, and then explain why the real number constructed in the proof is guaranteed not to be on the list of real numbers.

4.9 The Schröder-Bernstein Theorem

Theorem 4.8.3 shows that '$\leq$', as applied to infinite cardinal numbers, has some familiar properties, that is, some properties of '$\leq$' in more familiar settings, like the integers. Another property that we rely on when dealing with $\mathbb{Z}$ or $\mathbb{Q}$ or $\mathbb{R}$ is **anti-symmetry**: if $x \leq y$ and $y \leq x$ then $x = y$. It is far from obvious that the ordering of the infinite cardinals obeys this rule, but it does.

THEOREM 4.9.1 Schröder-Bernstein Theorem If $\overline{A} \leq \overline{B}$ and $\overline{B} \leq \overline{A}$, then $\overline{A} = \overline{B}$.

Proof. We may assume that A and B are disjoint sets. Suppose $f: A \to B$ and $g: B \to A$ are both injections; we need to find a bijection $h: A \to B$. Observe that if a is in A, there is at most one b_1 in B such that $g(b_1) = a$. There is, in turn, at most one a_1 in A such that $f(a_1) = b_1$. Continuing in this way, we can find a string of "ancestors" of a:

$$a = a_0, b_1, a_1, b_2, a_2, b_3, a_3, \ldots$$

such that $g(b_n) = a_{n-1}$ and $f(a_n) = b_n$. Call this the **lineage** of a. Of course, any $b \in B$ also has a lineage. Note that the lineage of $a \in A$ consists of just itself if a is not in the image of g; likewise, an element $b \in B$ might have no ancestors other than itself.

The lineage may take three forms: it may be infinite; it may end at some term a_k or b_k, if a_k is not in the image of g or b_k is not in the image of f; or it may "wrap around" to the beginning, if $a_k = a$ for some $k > 0$. If a lineage ends with a term a_k, $k \geq 0$, we say it **ends in** A. Let A_A and B_A be the subsets of A and B, respectively, consisting of those elements whose lineage ends in A.

Claim 1. If $f(a) = b$, then $a \in A_A$ iff $b \in B_A$. To see this, observe that the lineage of b is

$$b, a, b_1, a_1, b_2, a_2, b_3, a_3, \ldots$$

i.e., to get the lineage of b, just add it to the lineage of a. Now it is clear that if the lineage of a ends in A, so does the lineage of b. Suppose that the lineage of b ends in A. The lineage of b must then include a, and so the lineage of a ends in A also.

Now define $\hat{f}\colon A_A \to B_A$ by $\hat{f}(a) = f(a)$ (i.e., $\hat{f}(a)$ is f restricted to A_A, and with a different codomain.).

Claim 2. $\hat{f}$ is a bijection. Since f is an injection, it follows easily that $\hat{f}$ is an injection. To show $\hat{f}$ is surjective, suppose $b \in B_A$. Since the lineage of b ends in A, b must be in the image of f. So there is an $a \in A$ such that $f(a) = b$. Since $b \in B_A$, by claim 1, $a \in A_A$. Therefore, $\hat{f}(a) = b$ *for some* a *in* A_A, and $\hat{f}$ is surjective.

We outline a parallel construction and leave the details for the exercises. ${A_A}^c$ (the complement of A_A in A) and ${B_A}^c$ consist of those elements whose lineage does not end in A.

Claim $1'$. If $g(b) = a$, then $b \in {B_A}^c$ iff $a \in {A_A}^c$ (exercise 4).

Claim $1'$ allows us to define $\hat{g}\colon {B_A}^c \to {A_A}^c$, where $\hat{g}(b) = g(b)$ for any $b \in {B_A}^c$.

Claim $2'$. $\hat{g}$ is a bijection (exercise 5).

The theorem follows from claims 2 and $2'$: define $h\colon A \to B$ by the formula,

$$h(a) = \begin{cases} \hat{f}(a); & \text{if } a \in A_A, \\ \hat{g}^{-1}(a); & \text{if } a \in {A_A}^c. \end{cases}$$

It is straightforward to verify that h is a bijection (exercise 6). ∎

It is sometimes tempting to react to a result like this with, "Of course! How could it be otherwise?" This may be due in part to the use of the familiar symbol '$\leq$'—but of course, just using the symbol hardly guarantees that it acts like '$\leq$' in more familiar

contexts. Even paying attention to the new meaning, this theorem may seem "obvious." Perhaps the best way to see that it might not be so obvious is to look at a special case, one in which the injections f and g are easy to find, but there does not seem to be any "obvious" bijection. See exercise 8.

EXAMPLE 4.9.2 Suppose $D = \{\,(x, y) : x^2 + y^2 \leq 1\,\}$ is the unit disc in the plane and S is the square $\{\,(x, y) : -1 \leq x, y \leq 1\,\}$. Since $D \subseteq S$, $\overline{D} \leq \overline{S}$. The map $f\big((x, y)\big) = (x/2, y/2)$ is an injection $S \to D$, so $\overline{S} \leq \overline{D}$. By the Schröder-Bernstein Theorem, $\overline{S} = \overline{D}$. (So it is possible, after all, to fit a square peg in a round hole!) □

Felix Bernstein. Bernstein (1878–1956) studied under Cantor in Halle, and under Hilbert and Klein in Göttingen. It was in 1895 or 1896, while an undergraduate, that he proved the equivalence theorem for sets. Cantor had been working on the problem, but left for a holiday. In his absence, Bernstein was proof-reading one of Cantor's books; the idea for his proof of the equivalence theorem came to him one morning while he was shaving. Cantor later worked for several years to refine the proof to his satisfaction, but always gave full credit for the theorem to Bernstein.

After taking his undergraduate degree, Bernstein went to Pisa to study art. He was persuaded by two professors there to return to mathematics, after they heard Cantor lecture on the equivalence theorem. Bernstein remained interested in the arts, especially sculpture and painting, for the rest of his life.

Bernstein received his Ph.D. at Göttingen in 1901, and after some time in Halle became associate professor of mathematical statistics at Göttingen.

Bernstein was a versatile mathematician, working in both pure and applied mathematics. He was one of the first mathematicians to apply set theory to other branches of mathematics. By the 1920s, he had become interested in mathematical genetics, and made important contributions in population genetics. Most notable was his successful explanation of the inheritance of blood type, based on a set of three alleles.

In the 1930s, Bernstein emigrated to the United States, and became a citizen in 1940. He taught at Columbia, NYU and Syracuse until 1948, when he returned to Göttingen.

The information here is from the article on Bernstein, by Henry Nathan, in *Biographical Dictionary of Mathematicians*, New York: Charles Scribner's Sons, 1991.

Exercises 4.9.

1. Why is the Schröder-Bernstein Theorem easy if A and B are finite sets?
2. Suppose $f: A \to B$ is an injection and $g: A \to B$ is a surjection. Show there is a bijection $h: A \to B$.
3. At the beginning of the proof of the Schröder-Bernstein Theorem we said, "We may assume that A and B are disjoint sets." Why? In other words, what do we do if A and B are not disjoint?
4. Prove Claim $1'$ of theorem 4.9.1.
5. Prove Claim $2'$ of theorem 4.9.1.
6. Prove that the function h defined at the end of the proof of the Schröder-Bernstein Theorem is a bijection.
7. Use the Schröder-Bernstein Theorem to conclude that $\overline{[0,1]} = c$. (See exercise 3 of section 4.7.)
8. Find simple injections from $[0,1]$ to $\mathbb{R}$ and from $\mathbb{R}$ to $[0,1]$. Then find an explicit bijection from $[0,1]$ to $\mathbb{R}$.
9. Note that if A has m elements and B has n elements, then $A \times B$ has mn elements. We use this to define the product of two cardinals by the formula $\overline{A} \cdot \overline{B} = \overline{A \times B}$. Show that this definition is independent of the sets A and B, i.e., if $A \approx A'$ and $B \approx B'$, then $A \times B \approx A' \times B'$.
10. Show that $\aleph_0 \cdot \aleph_0 = \aleph_0$. (Hint: exercise 6 of section 4.7.) (See exercise 9.)
11. Show that if $\overline{A} \leq \overline{B}$ then $\overline{A} \cdot \overline{C} \leq \overline{B} \cdot \overline{C}$. (See exercise 9.)
12. Prove that if $\overline{A} < \overline{B} \leq \overline{C}$ then $\overline{A} < \overline{C}$.

4.10 Cantor's Theorem

We have now seen infinite sets of two different sizes, $\aleph_0$ and c. Are there others? Is there a largest infinite size, i.e., a largest cardinal number? Recall that for any set A, the **power set** of A, written $\mathcal{P}(A)$, is the collection of all subsets of A. For example, $\mathcal{P}(\{1,2\}) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. For finite sets, the power set is not just larger than the original set, it is *much* larger (see exercise 1). This makes it natural to hope that the power set of an infinite set will be larger than the base set.

Let $\overline{A} < \overline{B}$ mean that $\overline{A} \leq \overline{B}$, but A and B do not have the same cardinality. The next theorem answers both questions posed above.

THEOREM 4.10.1 Cantor's Theorem If A is any set, then $\overline{A} < \overline{\mathcal{P}(A)}$.

Proof. First, we need to show that $\overline{A} \leq \overline{\mathcal{P}(A)}$: define an injection $f: A \to \mathcal{P}(A)$ by $f(a) = \{a\}$. Now we need to show that there is no bijection $g: A \to \mathcal{P}(A)$. For a contradiction, suppose g is such a bijection. Let

$$S = \{a \in A : a \notin g(a)\} \subseteq A.$$

Since $S \in \mathcal{P}(A)$, $S = g(x)$, for some $x \in A$, because g is a surjection. There are two possibilities: $x \in S$ and $x \notin S$.

1. If $x \in S$, then $x \notin g(x) = S$, i.e., $x \notin S$, a contradiction.

2. If $x \notin S$, then $x \in g(x) = S$, i.e., $x \in S$, a contradiction.

Therefore, no such bijection is possible. ■

Cantor's theorem implies that there are infinitely many infinite cardinal numbers, and that there is no largest cardinal number. It also has the following interesting consequence:

There is no such thing as the "set of all sets".

Suppose A were the set of all sets. Since every element of $\mathcal{P}(A)$ is a set, we would have $\mathcal{P}(A) \subseteq A$, so

$$\overline{\mathcal{P}(A)} \leq \overline{A} \leq \overline{\mathcal{P}(A)}.$$

By the Schröder–Bernstein Theorem, $\overline{\mathcal{P}(A)} = \overline{A}$, but this contradicts Cantor's Theorem.

Many questions about the cardinal numbers remain. Since we know that $\mathbb{Z}$ and $\mathbb{Q}$ are the same size, and that $\mathbb{R}$ is larger, one very natural question is whether there are any sets 'between' $\mathbb{Z}$ and $\mathbb{R}$, that is, strictly bigger than $\mathbb{Z}$ (and $\mathbb{Q}$) but strictly smaller than $\mathbb{R}$. The **continuum hypothesis** says:

There is no set A with $\aleph_0 < \overline{A} < c$.

That is, the continuum hypothesis asserts that c is the first cardinal number larger than $\aleph_0$. Remarkably, the continuum hypothesis *cannot be proved to be true and cannot be proved to be false.* In the 1920's, Kurt Gödel showed that the continuum hypothesis cannot be *disproved*, and in the early 1960's, Paul Cohen showed that it cannot be *proved* either.

Georg Cantor. Cantor (1845–1918) was born in St. Petersburg and grew up in Germany. He took an early interest in theological arguments about continuity and the infinite, and as a result studied philosophy, mathematics and physics at universities in Zurich, Göttingen and Berlin, though his father encouraged him to pursue engineering. He did his doctorate in number theory and then worked in analysis before doing his pioneering work in the theory of sets.

The prevailing opinion in the nineteenth century was that 'completed' infinities could not be studied rigorously; only 'potential' infinity made sense—for example, the process of repeatedly adding one, starting at 1, would never finish and was therefore infinite, but most

mathematicians viewed the completed set of positive integers (or any other infinite set) as a dubious concept at best. An infinite set can be placed in one to one correspondence with a proper subset of itself; most mathematicians saw this as a paradox, and 'solved' the problem by declaring that 'infinite sets' simply make no sense.

A few mathematicians went against the grain; Dedekind realized that the 'paradoxical' correspondence between a set and one of its proper subsets could be taken as the *definition* of an infinite set. Cantor took this notion much further, showing that infinite sets come in an infinite number of sizes. Cantor knew most of what we have seen in this chapter: he showed that the rational numbers are countable, that $\mathbb{R}$ is not countable, that $\mathcal{P}(A)$ is always bigger than A. The **algebraic numbers** are those real numbers that are roots of polynomials with rational coefficients—for example, $\sqrt{2}$ is a solution of $x^2 - 2 = 0$, and is therefore irrational and algebraic (all rational numbers are algebraic). There are 'more' algebraic numbers than rational numbers, in the sense that the algebraic numbers form a proper superset of the rationals, but Cantor showed that the set of algebraic numbers is countable. This means that the **transcendental** numbers (that is, the non-algebraic numbers, like π and e) form an uncountable set—so in fact almost all real numbers are transcendental.

In addition to the arithmetic of infinite cardinal numbers, Cantor developed the theory of infinite ordinal numbers. The two concepts are practically the same for finite numbers, so the idea that infinite ordinals and infinite cardinals are different takes some getting used to. Since there is essentially only one way to make a total order out of four objects (namely, pick a first, a second, a third and a fourth), the cardinal number 4 ('how many') and the ordinal number 4 ('what order') are easily confused. For infinite sets the situation is radically different. The **ordinal** number of the positive integers, called ω, is simply the usual total ordering of the positive integers. 'Addition' of ordinals is accomplished by placing the orders side by side: $1 + \omega$ 'looks like' one item followed by a countable number of items *in the same order as the positive integers*—this looks just like the positive integers. On the other hand, $\omega + 1$ looks like the positive integers followed by a single item, and is much different than the usual ordering of the positive integers, even though the size of the two ordered sets is the same. (The easiest way to see that there is a crucial difference between the two orderings is to note that one element of $\omega + 1$ has an infinite number of predecessors, while all of the elements of $1 + \omega$ have a finite number of predecessors.)

Cantor was unable to secure a position at a major university, including Berlin, where he most desired to be. This failure was due in large part to the influence of Kronecker, a mathematician at Berlin, who ridiculed all talk of completed infinities, convinced that only finite processes could be justified. (As a result, he didn't believe in irrational numbers, since they could not be 'produced' by a finite process.) Beginning in 1884, Cantor suffered a series of nervous breakdowns, presumably related to the refusal of so many mathematicians

to accept his work; Cantor himself had occasional doubts about his results—the proofs were clear and rigorous, but the results still seemed paradoxical. Cantor died in a mental institution in 1918, though he did get some positive recognition for his work before his death. Writing a few years after Cantor's death, the great mathematician David Hilbert called Cantor's work "the most astonishing product of mathematical thought, one of the most beautiful realizations of human activity in the domain of the purely intelligible." The years since have more than justified this assessment of Cantor's work.

The information here is taken from *A History of Mathematics*, by Carl Boyer, New York: John Wiley & Sons, 1968. For a more detailed account of Cantor's life and work, see *Georg Cantor, His Mathematics and Philosophy of the Infinite*, by Joseph Dauben, Harvard University Press, 1979.

Exercises 4.10.

1. Verify Cantor's Theorem for finite sets by showing that if A has n elements, then $\mathcal{P}(A)$ has 2^n elements.

The representation of a real number as a decimal is almost, but not quite, unique. The problem arises only with those numbers that have "terminating" decimal expansions, like $1 = .99999\ldots$ and $.246 = .2459999\ldots$. A similar statement, of course, is true if we use some other base b. For example, in base 2, $1 = .11111\ldots$ and $.11 = .101111\ldots$.

Recall from exercise 7 of section 4.9 that $\overline{[0,1]} = c$. If b is a base for a number system, define a function $f_b\colon \mathcal{P}(\mathbb{N}) \to [0,1]$ as follows: if $S \subseteq \mathbb{N}$, let

$$f_b(S) = \sum_{i \in S} b^{-i}.$$

For example, writing expansions in base b,

$$\begin{aligned} f_b(\{1,2,3\}) &= .111000\ldots, \\ f_b(\{\text{odd numbers}\}) &= .10101010\ldots, \\ f_b(\{\text{prime numbers}\}) &= .011010100\ldots \end{aligned}$$

2. What kind of function is f_{10}? How about f_2?

3. Use exercise 2 to prove $\overline{\mathcal{P}(\mathbb{N})} = c$. Knowing this, the continuum hypothesis can be rephrased: There is no set A such that $\overline{\mathbb{N}} < \overline{A} < \overline{\mathcal{P}(\mathbb{N})}$.

4. Suppose A and B are sets.

a) If $\overline{A} \leq \overline{B}$, prove $\overline{\mathcal{P}(A)} \leq \overline{\mathcal{P}(B)}$.

b) Use part (a) to prove that if $\overline{A} = \overline{B}$, then $\overline{\mathcal{P}(A)} = \overline{\mathcal{P}(B)}$.

5. Note that if A and B are **disjoint** (i.e., $A \cap B = \emptyset$) finite sets with m and n elements, respectively, then $A \cup B$ has $m + n$ elements. We want to use this idea to define the sum of infinite cardinal numbers.

a) Suppose that A and B are sets, not necessarily disjoint. Show that $A_0 = A \times \{0\}$ and $B_1 = B \times \{1\}$ are disjoint, and show that $A \approx A_0$ and $B \approx B_1$. We use this to define the sum of two cardinal numbers by the formula $\overline{A} + \overline{B} = \overline{A_0 \cup B_1}$.

b) Show that this definition is independent of the sets A and B, i.e., if $A \approx A'$ and $B \approx B'$, then $A_0 \cup B_1 \approx A'_0 \cup B'_1$. (Find a bijection from $A_0 \cup B_1$ to $A'_0 \cup B'_1$.)

6. Use exercise 5 of section 4.8 to show that $\aleph_0 + \aleph_0 = \aleph_0$.

If A and B are sets, let $\text{map}(A, B)$ denote the collection of all functions from A to B. Observe that if A has n elements and B has m elements, $\text{map}(A, B)$ has m^n elements. We want to define $\overline{B}^{\overline{A}}$ to mean $\overline{\text{map}(A, B)}$. In order to do so, we need to verify that if $f\colon A \to A'$ and $g\colon B \to B'$ are bijections, we can find a bijection $h\colon \text{map}(A, B) \to \text{map}(A', B')$. To this end, if $\phi \in \text{map}(A, B)$, let $h(\phi) = g \circ \phi \circ f^{-1}$. Conversely, if $\phi \in \text{map}(A', B')$, let $k(\phi) = g^{-1} \circ \phi \circ f$.

7. Verify that h and k are inverse functions (and hence bijective).

If $S \subseteq A$, define the **characteristic function** of S by the equation,

$$\chi_S(x) = \begin{cases} 1, & \text{if } x \in S; \\ 0, & \text{if } x \notin S. \end{cases}$$

8. Show that associating S with χ_S defines a one-to-one correspondence between $\mathcal{P}(A)$ and $\text{map}(A, \{0, 1\})$. This implies that $\overline{\mathcal{P}(A)} = 2^{\overline{A}}$. (Hint: if $\phi \in \text{map}(A, \{0, 1\})$, for which $S \subseteq A$ does $\phi = \chi_S$?)

9. Suppose that a/b is a rational number. Show that it is algebraic by finding a polynomial with rational coefficients that has a/b as a root. Also, find a polynomial with integer coefficients that has a/b as a root.

5

Relations

We might arguably say that mathematics is the study of how various entities are related; in any case, the relationships between mathematical objects is a large part of what we study. You are already familiar with many such relationships: If $f(x) = y$ then x and y are related in a special way; if we say $x < y$ or $x = y$ or $x \geq y$, we are highlighting a particular relationship between x and y.

Certain kinds of relationships appear over and over in mathematics, and deserve careful treatment and study. We use the notation $x \sim y$ to mean that x and y are related in some special way; "$\sim$" is called a **relation**. The meaning of $\sim$ changes with context—it is not a fixed relation. In some cases, of course, we can use other symbols that have come to be associated with particular relations, like "$<$" and "$=$". We could give a formal definition of the term **relation**, but for our purposes an intuitive approach will be sufficient, just as we have made do without a formal definition of "function".

5.1 Equivalence Relations

We say $\sim$ is an **equivalence relation** on a set A if it satisfies the following three properties:

a) **reflexivity**: for all $a \in A$, $a \sim a$.

b) **symmetry**: for all $a, b \in A$, if $a \sim b$ then $b \sim a$.

c) **transitivity**: for all $a, b, c \in A$, if $a \sim b$ and $b \sim c$ then $a \sim c$.

EXAMPLE 5.1.1 Equality (=) is an equivalence relation. It is of course enormously important, but is not a very interesting example, since no two distinct objects are related by equality. □

EXAMPLE 5.1.2 Suppose A is $\mathbb{Z}$ and n is a fixed positive integer. Let $a \sim b$ mean that $a \equiv b \pmod{n}$. The fact that this is an equivalence relation follows from standard properties of congruence (see theorem 3.1.3). □

EXAMPLE 5.1.3 Let A be the set of all words. If $a, b \in A$, define $a \sim b$ to mean that a and b have the same number of letters; $\sim$ is an equivalence relation. □

EXAMPLE 5.1.4 Let A be the set of all vectors in $\mathbb{R}^2$. If $a, b \in A$, define $a \sim b$ to mean that a and b have the same length; $\sim$ is an equivalence relation. □

If $\sim$ is an equivalence relation defined on the set A and $a \in A$, let

$$[a] = \{x \in A : a \sim x\},$$

called the **equivalence class corresponding to** a. Observe that reflexivity implies that $a \in [a]$.

EXAMPLE 5.1.5 If A is $\mathbb{Z}$ and $\sim$ is congruence modulo 6, then

$$[2] = \{\ldots, -10, -4, 2, 8, \ldots\}.$$

□

EXAMPLE 5.1.6 Using the relation of example 5.1.3, [math] is the set consisting of all 4 letter words. □

EXAMPLE 5.1.7 Using the relation of example 5.1.4, $[(1, 0)]$ is the unit circle. □

THEOREM 5.1.8 Suppose $\sim$ is an equivalence relation on the set A. Then for all $a, b \in A$, the following are equivalent:

a) $a \sim b$,

b) $[a] \cap [b] \neq \emptyset$,

c) $[a] = [b]$.

Proof. (a) $\Rightarrow$ (b). Suppose $a \sim b$. Then b is an element of $[a]$. Since b is also in $[b]$, $b \in [a] \cap [b]$, so $[a] \cap [b] \neq \emptyset$.

(b) ⇒ (c). Suppose $y \in [a] \cap [b]$, that is, $a \sim y$ and $b \sim y$. We need to show that the two sets $[a]$ and $[b]$ are equal. If $x \in [a]$, then $b \sim y$, $y \sim a$ and $a \sim x$, so that $b \sim x$, that is, $x \in [b]$. Conversely, if $x \in [b]$, then $a \sim y$, $y \sim b$ and $b \sim x$, so that $a \sim x$, that is, $x \in [a]$.

(c) ⇒ (a). If $[a] = [b]$, then since $b \in [b]$, we have $b \in [a]$, that is, $a \sim b$. ■

Let $A/\sim$ denote the collection of equivalence classes; $A/\sim$ is a partition of A. (Recall that a partition is a collection of disjoint subsets of A whose union is all of A.) The expression "$A/\sim$" is usually pronounced "A mod twiddle."

EXAMPLE 5.1.9 Using the relation of example 5.1.5,

$$A/\sim = \{[0], [1], [2], [3], [4], [5]\} = \mathbb{Z}_6$$

□

EXAMPLE 5.1.10 Using the relation of example 5.1.3,

$$A/\sim = \{\{\text{one letter words}\}, \{\text{two letter words}\}, \{\text{three letter words}\}, \ldots\}$$

□

EXAMPLE 5.1.11 Using the relation of example 5.1.4, $A/\sim = \{C_r : 0 \leq r \in \mathbb{R}\}$, where for each $r > 0$, C_r is the circle of radius r centered at the origin and $C_0 = \{(0, 0)\}$. □

Exercises 5.1.

1. Suppose A is $\mathbb{Z}$ and n is a fixed positive integer. Let $a \sim b$ mean that $a \equiv b \pmod{n}$. Prove that $\sim$ is an equivalence relation.
2. Let $A = \mathbb{R}^3$. Let $a \sim b$ mean that a and b have the same z coordinate. Show $\sim$ is an equivalence relation and describe $[a]$ geometrically.
3. Suppose n is a positive integer and $A = \mathbb{Z}_n$. Let $a \sim b$ mean there is an element $x \in \mathbb{U}_n$ such that $ax = b$. Show $\sim$ is an equivalence relation. Compute the equivalence classes when $n = 12$.
4. Recall from section 3.9 the set $G_e = \{x \mid 0 \leq x < n, (x, n) = e\}$. There you find an example using $n = 12$, and the sets G_e bear a striking resemblence to the answer to the previous problem. For each divisor e of n, define $A_e = \{eu \bmod n \mid (u, n) = 1\}$, which are essentially the equivalence classes of the previous exercise. Prove that $A_e = G_e$.
5. Let S be some set and $A = \mathcal{P}(S)$. For any $a, b \in A$, let $a \sim b$ mean that a and b have the same cardinality. Show $\sim$ is an equivalence relation. Compute the equivalence classes when $S = \{1, 2, 3\}$.
6. The following purports to prove that the reflexivity condition is unnecessary, that is, it can be derived from symmetry and transitivity:

Suppose $a \sim b$. By symmetry, $b \sim a$. Since $a \sim b$ and $b \sim a$, by transitivity, $a \sim a$. Therefore, $\sim$ is reflexive.

What's wrong with this argument?

7. The example in 5.1.5 and 5.1.9 is a little peculiar, since at the time we defined $\mathbb{Z}_6$ we attached no "real" meaning to the notation $[x]$. Discuss.

8. Suppose $\sim$ is a relation on A that is reflexive and has the property that for all a, b, c, if $a \sim b$ and $a \sim c$, then $b \sim c$. Show $\sim$ is an equivalence relation.

9. Suppose $\sim_1$ and $\sim_2$ are equivalence relations on a set A. Let $\sim$ be defined by the condition that $a \sim b$ iff $a \sim_1 b \wedge a \sim_2 b$. Show $\sim$ is an equivalence relation on A. If $[a]$, $[a]_1$ and $[a]_2$ denote the equivalence class of a with respect to $\sim$, $\sim_1$ and $\sim_2$, show $[a] = [a]_1 \cap [a]_2$.

10. What happens if we try a construction similar to problem 9 with $\vee$ replacing $\wedge$?

11. Suppose $f: A \to B$ is a function and $\{Y_i\}_{i \in I}$ is a partition of B. Prove $\{f^{-1}(Y_i)\}_{i \in I}$ is a partition of A.

5.2 Factoring Functions

Suppose $f: A \to B$ is a function. For $x, y \in A$ let $x \sim y$ mean $f(x) = f(y)$.

LEMMA 5.2.1 $\sim$ is an equivalence relation on A.

Proof. Since $f(x) = f(x)$, $x \sim x$ and $\sim$ is reflexive. If $x \sim y$, then $f(x) = f(y)$, so $f(y) = f(x)$ and $y \sim x$; hence , $\sim$ is symmetric. If $x \sim y$ and $y \sim z$, then $f(x) = f(y)$ and $f(y) = f(z)$, so $f(x) = f(z)$, which implies that $x \sim z$, so $\sim$ is transitive. ■

EXAMPLE 5.2.2 Suppose $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $B = \{a, b, c, d, e\}$ and

$$\begin{array}{lll} f(1) = b, & f(4) = a, & f(7) = a, \\ f(2) = a, & f(5) = d, & f(8) = b, \\ f(3) = e, & f(6) = b, & f(9) = e. \end{array}$$

Then $A/\sim = \{\{1, 6, 8\}, \{2, 4, 7\}, \{3, 9\}, \{5\}\}$. □

EXAMPLE 5.2.3 Suppose $A = \mathbb{R}^2$, $B = \mathbb{R}$ and $f((x, y)) = x^2 + y^2$; then an equivalence class is a circle centered at the origin—for example, $[(1, 0)]$ is the unit circle. Thus, $A/\sim$ is the collection of all circles centered at the origin (if we call $\{(0, 0)\}$ a circle of radius 0). □

EXAMPLE 5.2.4 Suppose A is the set of all people, $B = \mathbb{Z}$, and $f: A \to B$ is the function that assigns to each person his or her age in years. An equivalence class consists of all people of some given age. □

Suppose C is the image of f. If $[x] \in A/\sim$, let $\overline{f}: (A/\sim) \to C$ be defined by $\overline{f}([x]) = f(x)$.

THEOREM 5.2.5 $\overline{f}$ is a bijection.

Proof. Note that $[x] = [y]$ iff $x \sim y$ iff $f(x) = f(y)$. Reading these in the direction $\Rightarrow$ shows that the definition of $\overline{f}$ does not depend on which representative of a given equivalence class is used, so $\overline{f}$ is well defined. Reading in the direction $\Leftarrow$ shows that $\overline{f}$ is an injection. Since the image of $\overline{f}$ is C, $\overline{f}$ is a bijection. ■

EXAMPLE 5.2.6 In 5.2.2, $C = \{a, b, d, e\}$, $\overline{f}(\{1, 6, 8\}) = b$, $\overline{f}(\{2, 4, 7\}) = a$, $\overline{f}(\{3, 9\}) = e$, $\overline{f}(\{5\}) = d$. □

EXAMPLE 5.2.7 In example 5.2.3, if $C_r \subseteq A$ is the circle of radius r, then $\overline{f}(C_r) = r^2$. □

EXAMPLE 5.2.8 In example 5.2.4, if P_n is the set of all people of age n, then $\overline{f}(P_n) = n$. □

DEFINITION 5.2.9 Let $\pi: A \to (A/\sim)$ be defined by $\pi(x) = [x]$, so $\pi(x) = \pi(y)$ if and only if $x \sim y$. Note that π is a surjection since every equivalence class has at least one representative. □

THEOREM 5.2.10 Every function is, in a natural way, the composition of an injection, a bijection and a surjection.

Proof. Given $f: A \to B$, with image C, let $g: C \to B$ be the inclusion map, an injection. We have already observed that $\overline{f}: (A/\sim) \to C$ is a bijection and $\pi: A \to A/\sim$ is a surjection. Now for any $x \in A$,

$$(g \circ \overline{f} \circ \pi)(x) = g(\overline{f}(\pi(x))) = g(\overline{f}([x])) = g(f(x)) = f(x).$$

So $f = g \circ \overline{f} \circ \pi$ as required. ■

Exercises 5.2.

1. Suppose $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, $B = \{a, b, c, d, e, f\}$ and

$$\begin{array}{lll} f(1) = b, & f(5) = f, & f(9) = b, \\ f(2) = a, & f(6) = e, & f(10) = b, \\ f(3) = e, & f(7) = a, & f(11) = a, \\ f(4) = f, & f(8) = b, & f(12) = a. \end{array}$$

Find $C = f(A)$, $A/\sim$, and $\overline{f}$.

2. Suppose $A = \mathbb{R}^3$, $B = \mathbb{R}^2$ and $f: A \to B$ is $f((x, y, z)) = (0, z)$. Describe $C = f(A)$, $A/\sim$, and $\overline{f}$. Use the relation $\sim$ defined at the beginning of the section.

3. When is π an injection? When is g (as defined in theorem 5.2.10) a surjection?

4. Suppose A and B have m and n elements, respectively, and the image, C, of $f: A \to B$ has k elements. If $A/\sim = \{S_1, S_2, \ldots, S_k\}$ and each S_i has m_i elements in it, how many pseudo-inverses does f have? When does f have exactly one pseudo-inverse?
5. If h is a pseudo-inverse to f, what is $\pi \circ h \circ g \circ \overline{f}$? ($g$ is as defined in theorem 5.2.10.)
6. If $X \subseteq A$, show that $f^{-1}(f(X)) = \bigcup_{x \in X}[x]$, using the relation $\sim$ defined at the beginning of the section.

5.3 Ordered Sets

If A is a set, then a relation $\leq$ on A is a *partial ordering* if

1) for all $x \in A$, $x \leq x$ ($\leq$ is reflexive),

2) for all $x, y, z \in A$, if $x \leq y$ and $y \leq z$, then $x \leq z$ ($\leq$ is transitive),

3) for all $x, y \in A$, if $x \leq y$ and $y \leq x$, then $x = y$, ($\leq$ is **anti-symmetric).**

WARNING: we are appropriating the familiar symbol "$\leq$" to mean something new. The usual orderings of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ denoted by $\leq$ are partial orderings, so this use is "backwards compatible," but there are many other partial orderings, as we will see.

EXAMPLE 5.3.1 If $A = \mathbb{N}$ and $x \leq y$ means $x|y$, then $\leq$ is a partial ordering, so $\mathbb{N}$ has more than one "natural" partial ordering defined on it. (Unless we say otherwise, we will continue to use $\leq$ to denote the usual order for $\mathbb{N}$.) □

EXAMPLE 5.3.2 Suppose S is a set and A is the set of all functions $f: S \to \mathbb{R}$. Let $f \leq g$ mean $f(x) \leq g(x)$ for all $x \in S$; $\leq$ is a partial ordering of A. Notice that we have used the symbol $\leq$ in two different ways! □

EXAMPLE 5.3.3 Suppose S is a set and $A = \mathcal{P}(S)$ is the power set of S. Let $X \leq Y$ mean $X \subseteq Y$; $\leq$ is a partial ordering. □

DEFINITION 5.3.4 If $\leq$ is a partial ordering on A, we say it is a **total ordering** if for all $x, y \in A$, either $x \leq y$ or $y \leq x$. □

EXAMPLE 5.3.5 The familiar partial orderings of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are total orderings. Those of examples 5.3.1, 5.3.2 and 5.3.3 are not. □

DEFINITION 5.3.6 If $\leq$ is a partial ordering on A, and $S \subseteq A$, we say $x \in S$ is a **least element of** S if $x \leq y$ for all $y \in S$. We say $x \in S$ is a **greatest element of** S if $y \leq x$ for all $y \in S$. □

THEOREM 5.3.7 If $\leq$ is a total ordering on A, then every non-empty finite subset S of A has a least element and a greatest element.

Proof. We show there is a least element and leave the rest to an exercise. By induction. Let $P(n)$ be the statement "every subset of A having n elements has a least element." Clearly $P(1)$ is true. Suppose $P(n)$ is true. If S has $n+1$ elements, let $S = S' \cup \{y\}$, for some $y \in S$ and S' with n elements (namely, $S' = S \backslash \{y\}$, or S with y removed). By induction, S' has a least element, call it x. Since $\leq$ is a total ordering, either $x \leq y$ or $y \leq x$. In the first case, x is a least element of S. On the other hand, if $y \leq x$ we claim that y is a least element of S: If $z \in S$, then either $z = y$, or $z \in S'$ and $y \leq x \leq z$. In either case, $y \leq z$, as desired. ■

DEFINITION 5.3.8 We say $\leq$ is a **well ordering of** A if every non-empty subset S of A has a least element. □

THEOREM 5.3.9 If $\leq$ is a well ordering of A then it is a total ordering.

Proof. Suppose $x, y \in A$. Let $S = \{x, y\} \neq \emptyset$. By hypothesis, S has a smallest element. If it is x, then $x \leq y$, and if it is y, then $y \leq x$. ■

EXAMPLE 5.3.10 Since the partial orderings of examples 5.3.1, 5.3.2 and 5.3.3 are not total orderings, they are not well orderings. Since there is no smallest integer, rational number or real number, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are not well ordered. □

The following important fact is called the **well ordering principle.**

THEOREM 5.3.11 The usual ordering of $\mathbb{N}$ is a well ordering.

Proof. Let S be a non-empty subset of $\mathbb{N}$. Choose $n \in S$, and let $S_n = \{s \in S : s \leq n\}$. S_n is not empty, since it contains n, and has a finite number of elements, so by theorem 5.3.7 it has a least element x. Then x is, in fact, a least element of S, since if $y \in S - S_n$, then $x \leq n \leq y$, so $x \leq y$. ■

If $\leq$ is a partial ordering of A, then $x < y$ means $x \leq y$ and $x \neq y$, $x \geq y$ means $y \leq x$ and $x > y$ means $y < x$. These have some familiar properties, explored in the exercises.

Exercises 5.3.

In the following exercises, assume $\leq$ is a partial ordering of a set A.

1. Show that if $a < b$ then it is not true that $b \leq a$.
2. Show that at most one of the three properties $a < b$, $a = b$, $b < a$ is true.
3. If $a < b$ and $b \leq c$, show $a < c$.

4. If $a \leq b$ and $b < c$, show $a < c$.
5. Suppose $S \subseteq A$ has a least element. Show that it is unique.
6. Show that A is totally ordered if and only if every non-empty finite subset of A has a least element.
7. If A is totally ordered but not well ordered, prove there is a sequence of elements such that $a_1 > a_2 > a_3 > \cdots$. (Hint: start with a non-empty subset that does not contain a least element.)
8. Suppose A has the property that every non-empty countable subset has a least element. Show that A is well ordered. (Hint: show first that A is totally ordered, then use problem 7.)
9. Finish the proof of theorem 5.3.7.

5.4 New Orders from Old

Suppose $\leq$ is a partial ordering of A and S is a subset of A. If we restrict $\leq$ to S, we have an ordering of S.

LEMMA 5.4.1 Suppose $\leq$ is a partial ordering of A and S is a subset of A. Then

a) $\leq$ is a partial ordering of S;

b) if A is totally ordered by $\leq$, so is S;

c) if A is well ordered by $\leq$, so is S.

Proof. If $x \in S$, then since $x \in A$, $x \leq x$, so $\leq$ is reflexive on S. Similarly, transitivity, anti-symmetry, total ordering or well ordering on S follow from the corresponding property on A. ■

EXAMPLE 5.4.2 Any collection of subsets of X forms a subset of $\mathcal{P}(X)$, so the subsets are partially ordered by inclusion. □

EXAMPLE 5.4.3 Any subset of the rational numbers is countable and totally ordered by the usual ordering of $\mathbb{Q}$. □

Suppose $\leq_1$ is a partial ordering of A and $\leq_2$ is a partial ordering of B. On $A \times B$, let $(a, b) \leq (x, y)$ mean $a <_1 x$, or $a = x$ and $b \leq_2 y$. This is called the *lexicographic ordering*; the name is derived from its similarity to ordering words alphabetically. Note that if $(a, b) \leq (x, y)$ then $a \leq_1 x$.

THEOREM 5.4.4 Suppose $\leq_1$ is a partial ordering of A, $\leq_2$ is a partial ordering of B and $\leq$ is the lexicographic ordering.

a) $\leq$ is a partial ordering of $A \times B$.

b) If $\leq_1$ and $\leq_2$ are total orderings, so is $\leq$.

c) If $\leq_1$ and $\leq_2$ are well orderings, so is $\leq$.

Proof. (a) Since $a \leq_1 a$ and $b \leq_2 b$, $(a, b) \leq (a, b)$, i.e., $\leq$ is reflexive. Transitivity is left as an exercise. To show anti-symmetry, suppose $(a, b) \leq (x, y)$ and $(x, y) \leq (a, b)$. Looking at the first coordinates, we have $a \leq_1 x$ and $x \leq_1 a$. Since $\leq_1$ is anti-symmetric, $a = x$; in particular, it is not true that $a <_1 x$ or $x <_1 a$. So looking at second coordinates, $b \leq_2 y$ and $y \leq_2 b$, so $b = y$, as desired.

We leave part (b) to the exercises. As to part (c), suppose S is a non-empty subset of $A \times B$. Let $T = \{a : \exists b \in B\ ((a, b) \in S)\}$. Since S is non-empty, so is T. Let a_0 be the least element of T. Let $U = \{b : (a_0, b) \in S\}$. Since $a_0 \in T$, U is non-empty. Let b_0 be the least element of U. We claim that (a_0, b_0) is the least element of S. If $(x, y) \in S$, then by the definition of T, $x \in T$, so $a_0 \leq_1 x$. If $a_0 <_1 x$, then $(a_0, b_0) \leq (x, y)$, as required. Otherwise, $a_0 = x$ and $(x, y) = (a_0, y)$. So $y \in U$, and by definition, $b_0 \leq y$. This means $(a_0, b_0) \leq (x, y)$, as required. ∎

COROLLARY 5.4.5 The lexicographic ordering on $\mathbb{N} \times \mathbb{N}$ is a well ordering.

Proof. Immediate by 5.3.11 and 5.4.4. ∎

Exercises 5.4.

1. If $A = \{a, b, c\}$ and $B = \{a, b, c, d\}$ are ordered alphabetically, write down the lexicographic ordering of $A \times B$ (i.e., order the elements from smallest to largest).
2. If A and B are partially ordered and a_0 and b_0 are the least elements of A and B, show that (a_0, b_0) is the least element of $A \times B$ in the lexicographic ordering.
3. Suppose A is partially ordered, and every proper subset of A is totally ordered. Show that if $|A| \geq 3$, then A is totally ordered. What about $|A| = 2$?
4. Suppose $A = B \cup C$ is totally ordered. Show that A is well ordered if and only if B and C are well ordered.
5. Suppose A and B are non-empty partially ordered sets and $A \times B$ (ordered lexicographically) is totally ordered. Show that A and B are totally ordered.
6. Suppose A and B are non-empty partially ordered sets and $A \times B$ (ordered lexicographically) is well ordered. Show that A and B are well ordered.
7. Prove transitivity in 5.4.4(a).
8. Prove 5.4.4(b).

5.5 Partial Orders and Power Sets

Suppose A and B are partially ordered sets. We use "$\leq$" to denote both orders, instead of the more cumbersome "$\leq_A$" and "$\leq_B$", but keep in mind that the two orders are (potentially) much different. A bijection $f: A \to B$ is called an **isomorphism** if for all

$x, y \in A$, $x \leq y$ if and only if $f(x) \leq f(y)$. If there is such a function we say A and B are **isomorphic**.

When two partial orders are isomorphic, then *as partial orders* they are, in an obvious sense, really *the same partial order.* Of course, A and B may have other properties that make them much different from each other.

EXAMPLE 5.5.1 Let $A = \mathcal{P}(\{0,1\})$, ordered by $\subseteq$. Let $B = \{1,2,3,6\}$, with $n \leq m$ if and only if $n|m$. Then A and B are isomorphic. □

THEOREM 5.5.2 Suppose A, B and C are partially ordered sets.

a) A is isomorphic to itself.

b) If A is isomorphic to B, then B is isomorphic to A.

c) If A is isomorphic to B and B is isomorphic to C, then A is isomorphic to C.

Proof. Part (a) is trivial (use the identity function). Part (b) we leave as an exercise. For part (c), suppose $f: A \to B$ and $g: B \to C$ are isomorphisms. Note that $g \circ f$ is a bijection, and if $x, y \in A$, then $x \leq y$ if and only if $f(x) \leq f(y)$ (since f is an isomorphism) if and only if $g(f(x)) \leq g(f(y))$ (since g is an isomorphism). ■

EXAMPLE 5.5.3 If a is any real number, let $I_a = (-\infty, a] \subseteq \mathbb{R}$, and let $\mathcal{S}$ be the collection of all of the intervals I_a, $\mathcal{S} = \{I_a : a \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R})$, ordered by inclusion. Define $\phi: \mathbb{R} \to \mathcal{S}$ by $\phi(a) = I_a$; ϕ is a bijection. Note that $a \leq b$ if and only if $(\infty, a] \subseteq (\infty, b]$, so ϕ is an isomorphism. □

We can generalize this example. Suppose $\leq$ is a partial ordering of a set A. If $a \in A$, let $I_a = \{x \in A : x \leq a\}$; we call this the *interval determined by* a (but notice that it doesn't "look like" an interval in the more familiar sense). Let $\mathcal{S} = \{I_a : a \in A\} \subseteq \mathcal{P}(A)$, ordered by inclusion. Define $\phi: A \to \mathcal{S}$ by $\phi(a) = I_a$; ϕ is an isomorphism, as we prove next.

THEOREM 5.5.4 Any partially ordered set is isomorphic to a subset of a power set, ordered by the subset relation.

Proof. Let ϕ be as above. We show first that ϕ is bijective. By the definition of $\mathcal{S}$, ϕ is surjective. To show that it is injective, suppose $a, b \in A$ and $\phi(a) = \phi(b)$. Since $a \leq a$, $a \in I_a = \phi(a) = \phi(b) = I_b$, so $a \leq b$. Similarly, $b \leq a$, so $a = b$, and ϕ is injective. Now, given $a, b \in A$, we need to show that $a \leq b$ if and only if $I_a \subseteq I_b$. Suppose first that $I_a \subseteq I_b$; then $a \in I_a \subseteq I_b$ implies that $a \leq b$. Conversely, suppose $a \leq b$; then for any $x \in I_a$, $x \leq a$ and $a \leq b$, so $x \leq b$ and hence $x \in I_b$. This shows that $I_a \subseteq I_b$, and finishes the proof. ■

EXAMPLE 5.5.5 Suppose for $a, b \in \mathbb{N}$, $a \leq b$ means $a|b$. Then $I_6 = \{1, 2, 3, 6\}$ and $I_{12} = \{1, 2, 3, 4, 6, 12\}$. Note that $6|12$ and $I_6 \subseteq I_{12}$. The theorem implies that for any $a, b \in \mathbb{N}$, a divides b if and only if the set of divisors of a is a subset of the set of divisors of b. You should be able to see that this is true directly, that is, without using the theorem. □

Exercises 5.5.

1. List two isomorphisms from A to B in example 5.5.1. Give a bijection from A to B that is *not* an isomorphism.
2. Let $A = \{1, 2, 3, 4, 5, 6\}$ using the natural ordering. Find $\mathcal{S}$ and the function $f: A \to \mathcal{S}$.
3. Let $A = \{1, 2, 3, 4, 6, 12\}$ ordered by divisibility. Find $\mathcal{S}$ and the function $f: A \to \mathcal{S}$.
4. Prove that for any $a, b \in \mathbb{N}$, a divides b if and only if the set of divisors of a is a subset of the set of divisors of b. (Prove this directly, without using theorem 5.5.4.)
5. Suppose $\mathcal{S}$ is a collection of sets ordered by inclusion. If $\bigcap_{A \in \mathcal{S}} A$ is a member of $\mathcal{S}$, show that it is the least element of $\mathcal{S}$.
6. Prove 5.5.2(b).
7. Suppose A and B are isomorphic and A is totally ordered. Prove that B is totally ordered.
8. Suppose A and B are isomorphic and A is well ordered. Prove that B is well ordered.
9. Show that $\mathbb{Q}$ and $\mathbb{Z}$ are not isomorphic. (Hint: think of "order properties" satisfied by one, but not the other.)

5.6 Countable total orders

The rational numbers $\mathbb{Q}$ are a countable, totally ordered set, so any subset of the rationals is also countable and totally ordered. In fact, the subsets of the rationals are the 'only' countable, totally ordered sets!

EXAMPLE 5.6.1 Let $A = \mathbb{N} \times \mathbb{N}$ using the lexicographic ordering. Under this ordering, A is totally ordered (in fact, well ordered). We show that A is isomorphic to a subset of $\mathbb{Q}$. Let $f: A \to \mathbb{Q}$ be the function

$$f((n, m)) = 2n - \frac{1}{m},$$

and let B be the image of f. Clearly $f: A \to B$ is surjective. To convince yourself that f is an isomorphism, look at some values:

$$\begin{array}{lll} f((1,1)) = 2 - 1, & f((1,2)) = 2 - 1/2, & f((1,3)) = 2 - 1/3, \quad \ldots, \\ f((2,1)) = 4 - 1, & f((2,2)) = 4 - 1/2, & f((2,3)) = 4 - 1/3, \quad \ldots, \\ f((3,1)) = 6 - 1, & f((3,2)) = 6 - 1/2, & f((3,3)) = 6 - 1/3, \quad \ldots, \end{array}$$

In general, if we fix n, then $f((n, m))$ is a sequence that increases from $2n - 1$ toward $2n$. These rational numbers are ordered just like the lexicographic ordering of $\mathbb{N} \times \mathbb{N}$. □

This example may seem surprising at first, because the rationals seem "one-dimensional" while $\mathbb{N} \times \mathbb{N}$ seems "two-dimensional."

THEOREM 5.6.2 Suppose A is any countable, totally ordered set. Then A is isomorphic to a subset of the rational numbers.

Proof. Since A is countable, we can arrange it in a sequence $a_1, a_2, a_3, \ldots$. We describe a procedure to define $f(a_i)$ for each a_i in turn. Let $f(a_1)$ be any rational. Suppose we have defined $f(a_1), f(a_2), \ldots, f(a_n)$ in such a way that all order relations are preserved (that is, for all $i, j \leq n$, $a_i \leq a_j$ if and only if $f(a_i) \leq f(a_j)$). We want to define f on $a_{n+1} \in A$. Partition the set $\{a_1, \ldots, a_n\}$ into two subsets:

$$X = \{a_i : i \leq n \text{ and } a_i < a_{n+1}\},\ Y = \{a_i : i \leq n \text{ and } a_i > a_{n+1}\}.$$

In $\mathbb{Q}$, every element of $f(X)$ is smaller than every element of $f(Y)$. Choose q strictly larger than the elements of $f(X)$ and strictly smaller than the elements of $f(Y)$. For each $i \leq n$, the relationship between q and $f(a_i)$ is the same as the relationship between a_{n+1} and a_i. Therefore, letting $f(a_{n+1}) = q$, we have extended the function to one more element in such a way that all order relations are preserved. The resulting function defined on all of A is thus an isomorphism from A to the range of f. ∎

Exercises 5.6.

1. Show that the positive rationals, $\mathbb{Q}^+$ are isomorphic to the negative rationals, $\mathbb{Q}^-$. (Hint: $-1/x$.)

2. Show that $\{0, 1\} \times \mathbb{Z}$ (using the natural ordering on each factor and the lexicographic ordering on the product) is isomorphic to

$$\{\ldots, -4, -2, -1, -1/2, -1/4, \ldots, 1/4, 1/2, 1, 2, 4, \ldots\}.$$

3. Let $I = \{q \in \mathbb{Q} : 0 \leq q < 1\}$. Show that $\mathbb{Z} \times I$ (with the lexicographic ordering) is isomorphic to $\mathbb{Q}$. (Hint: add.)

If A and B are partially ordered sets, then $f\colon A \to B$ is an **embedding** if for all $x, y \in A$, $x \leq y$ iff $f(x) \leq f(y)$.

4. Show that an embedding is necessarily an injection, and hence A is isomorphic to the image of f.

5. Verify that the identity function is an embedding and that the composition of two embeddings is an embedding.

Suppose A and B are partially ordered sets and there is an embedding of A in B and an embedding of B in A. We would like to conclude that A and B are isomorphic—sort of a Schröder-Bernstein theorem for partial orders.

6. Show that this is true if A is finite.

7. Show that this does not hold in general by finding two totally ordered sets that can be embedded in each other but are not isomorphic. (Hint: try intervals.)

In spite of this, it can be proved that when A and B are *well ordered* and each can be embedded in the other, then they are isomorphic.

Bibliography

American Council of Learned Societies. *Biographical Dictionary of Mathematicians.* Charles Scribner's Sons, New York, 1991.

Carl B. Boyer. *A History of Mathematics.* John Wiley and Sons, New York, 1968.

Joseph Dauben. *Georg Cantor, His Mathematics and Philosophy of the Infinite.* Harvard University Press, Cambridge, MA, 1979.

Howard Eves. *An Introduction to the History of Mathematics.* Holt, Rinehart and Winston, New York, 1976.

G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Oxford University Press, London, fourth edition, 1960.

William Judson LeVeque. *Topics in Number Theory.* Addison–Wesley, Reading, MA, 1956.

Alexander Macfarlane. *Lectures on Ten British Mathematicians of the Nineteenth Century.* John Wiley & Sons, New York, 1916.

Index

S

T

U

V

W